Clustering, host halos, and environment of z 2 galaxies as a function of their physical properties

Matthieu Béthermin, Martin Kilbinger, Emanuele Daddi, Jared Gabor, Alexis Finoguenov, Henry Mccracken, Melody Wolk, Hervé Aussel, Veronica Strazzulo, Emeric Le Floc'h, et al.

A&A 567, A103 (2014)
DOI: 10.1051/0004-6361/201423451

Astronomy & Astrophysics

Clustering, host halos, and environment of $z \sim 2$ galaxies as a function of their physical properties

Matthieu Béthermin[1,2], Martin Kilbinger[1], Emanuele Daddi[1], Jared Gabor[1], Alexis Finoguenov[3], Henry McCracken[4], Melody Wolk[4], Hervé Aussel[1], Veronica Strazzulo[1], Emeric Le Floc'h[1], Raphaël Gobat[1], Giulia Rodighiero[5], Mark Dickinson[6], Lingyu Wang[7], Dieter Lutz[8], and Sébastien Heinis[9]

[1] Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu – CNRS – Université Paris Diderot, CEA-Saclay, pt courrier 131, 91191 Gif-sur-Yvette, France
e-mail: matthieu.bethermin@laposte.net
[2] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
e-mail: matthieu.bethermin@eso.org
[3] Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2a, 00014 Helsinki, Finland
[4] Institut d'Astrophysique de Paris, UMR 7095 CNRS, Université Pierre et Marie Curie, 98 bis boulevard Arago, 75014 Paris, France
[5] Dipartimento di Astronomia, Universita di Padova, Vicolo dell'Osservatorio 3, Padova 35122, Italy
[6] National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson AZ 85719, USA
[7] Computational Cosmology, Department of Physics, University of Durham, South Road, Durham, DH1 3LE, UK
[8] MPE, Postfach 1312, 85741 Garching, Germany
[9] Department of Astronomy, University of Maryland, College Park, MD 20742-2421, USA

Received 17 January 2014 / Accepted 27 April 2014

ABSTRACT

Using a sample of 25 683 star-forming and 2821 passive galaxies at $z \sim 2$, selected in the COSMOS field following the BzK color criterion, we study the hosting halo mass and environment of galaxies as a function of their physical properties. *Spitzer* and *Herschel* allow us to obtain accurate star formation rate estimates for starburst galaxies. We measure the autocorrelation and cross-correlation functions of various galaxy subsamples and infer the properties of their hosting halos using both a halo occupation model and the linear bias at large scale. We find that passive and star-forming galaxies obey a similarly rising relation between the halo and stellar mass. The mean host halo mass of star-forming galaxies increases with the star formation rate between 30 $M_\odot\,\mathrm{yr}^{-1}$ and 200 $M_\odot\,\mathrm{yr}^{-1}$, but flattens for higher values, except if we select only main-sequence galaxies. This reflects the expected transition from a regime of secular coevolution of the halos and the galaxies to a regime of episodic starburst. We find similar large-scale biases for main-sequence, passive, and starburst galaxies at equal stellar mass, suggesting that these populations live in halos of the same mass. However, we detect an excess of clustering on small scales for passive galaxies and showed, by measuring the large-scale bias of close pairs of passive galaxies, that this excess is caused by a small fraction (∼16%) of passive galaxies being hosted by massive halos ($\sim 3 \times 10^{13}\ M_\odot$) as satellites. Finally, extrapolating the growth of halos hosting the $z \sim 2$ population, we show that $M_\star \sim 10^{10}\ M_\odot$ galaxies at $z \sim 2$ will evolve, on average, into massive ($M_\star \sim 10^{11}\ M_\odot$), field galaxies in the local Universe and $M_\star \sim 10^{11}\ M_\odot$ galaxies at $z = 2$ into local, massive, group galaxies. We also identify two $z \sim 2$ populations which should end up in today's clusters: massive ($>M_\star \sim 10^{11}\ M_\odot$), strongly star-forming (>200 $M_\odot\,\mathrm{yr}^{-1}$), main-sequence galaxies, and close pairs of massive, passive galaxies.

Key words. galaxies: statistics – galaxies: halos – galaxies: formation – galaxies: evolution – infrared: galaxies – galaxies: starburst

1. Introduction

Understanding galaxy formation and evolution in the context of the standard ΛCDM cosmological model is one of the main challenges of modern astrophysics. For almost two decades, semi-analytical models have tried to reproduce the statistical properties of galaxies from the evolution of dark matter structures and using analytical recipes for baryonic physics calibrated on hydrodynamical simulations (e.g., Guiderdoni et al. 1998; Somerville & Primack 1999; Hatton et al. 2003). However, these models are not able to accurately reproduce the infrared and sub-millimeter number counts of galaxies, which directly probe star formation in galaxies at high redshift, without invoking strong assumptions like advocating the adoption of a top-heavy initial mass function (IMF) in high redshift major and minor mergers (Baugh et al. 2005), ad hoc inefficiencies of star-formation in low-mass halos (Bouché et al. 2010; Cousin et al. 2013), or excessively long delays for the re-accretion of the gas ejected by supernovae (Henriques et al. 2013). This population of high-redshift ($z > 2$), intensely star-forming ($SFR > 200\ M_\odot$/yr) galaxies is important, because these galaxies are thought to be the progenitors of massive and passive galaxies present in the local Universe (e.g., Daddi et al. 2007; Tacconi et al. 2008; Cimatti et al. 2008). The exact nature of the mechanism(s), that trigger the transformation of star-forming galaxies into passive elliptical galaxies is also an open question. Feedback from active galactic nuclei (AGNs) is often advocated for this (e.g., Cattaneo et al. 2006; Somerville et al. 2008), but other mechanisms such as the suppression of gas cooling in the most massive halos due to their hot atmosphere (e.g., Kereš et al. 2005; Birnboim et al. 2007) are also possible.

Recent observational studies revealed interesting insights about the nature of star formation in high-redshift massive galaxies. Measurements of the stellar mass function of galaxies

showed that a significant fraction (∼30%) of massive galaxies ($M_\star > 10^{11}\ M_\odot$) are already passive at $z \sim 2$, when the vast majority of low-mass galaxies are star-forming (Ilbert et al. 2010, 2013; Muzzin et al. 2013). In addition, detailed studies of star-forming galaxies at these redshifts found a strong correlation between the star formation rates (SFRs) and stellar masses (e.g., Daddi et al. 2007; Rodighiero et al. 2010), the so-called main sequence, which suggests that the star-formation is driven by universal, secular processes. *Herschel* showed that a small percent of the massive star-forming galaxies are strong outliers of this sequence and present an excess of specific SFR ($sSFR = SFR/M_\star$) by a factor of at least four compared to the main sequence (Elbaz et al. 2011; Rodighiero et al. 2011; Sargent et al. 2012). These episodic starbursts are probably induced by major mergers (Daddi et al. 2010; Hung et al. 2013). However, the strong diversity of the star-formation properties in galaxies at high redshift is not well understood. In principle, it could be related to the properties of the host dark matter halos, hence to environmental effects. This can be investigated by measuring the clustering of galaxies of different types. For instance, the clustering of high-redshift starbursts can discriminate between major-merger-driven and secular star-formation processes (van Kampen et al. 2005).

Because of the invisible nature of the dark matter, measuring the hosting halo mass of a galaxy sample is difficult. Nevertheless, the link between halo mass (M_h) and stellar mass ($M_\star$) was studied extensively during the last decade. The halo occupation modeling allows us to infer how galaxies are distributed inside halos from observations of their clustering (Cooray & Sheth 2002, for a review). Using this technique, Coupon et al. (2012) measured the relation between halo and stellar mass up to $z = 1.2$. A slightly different but complementary approach is the abundance matching technique, which connects halo mass and stellar mass of galaxies directly from the related mass functions, assuming a monotonic relation between these two quantities (e.g., Vale & Ostriker 2004). Finally, weak gravitational lensing can also provide strong constraints on the characteristic halo mass hosting a galaxy population (e.g., Mandelbaum et al. 2006). Leauthaud et al. (2012) made a combined analysis of $z < 1$ galaxies combining all these techniques and strongly constrained the $M_\star - M_\mathrm{h}$ relation. However, only a few studies extended these results at higher redshift. Among these, there are studies based on abundance matching going up to $z = 4$ by Behroozi et al. (2010) and Moster et al. (2010) and a work based on abundance matching and clustering at $z < 2$ by Wang et al. (2013). Recently, Wolk et al. (in prep.) pushed the studies of the $M_\star - M_\mathrm{h}$ relation up to $z \sim 2.5$ and lower stellar masses using both clustering and abundance matching with the UltraVISTA data (McCracken et al. 2012).

The link between halo mass and other properties like the SFR or sSFR has been explored less. However, some interesting analyses were recently performed. Empirical models (e.g., Conroy & Wechsler 2009; Behroozi et al. 2013) calibrated on the evolution of stellar mass function applied a simple prescription to estimate the mean relation between M_h and SFR. Other empirical models used the link between $M_\star$ and SFR estimated from UV and far-infrared observations and the well-studied $M_\mathrm{h} - M_\star$ relation to determine in turn the link between SFR and M_h (e.g., Béthermin et al. 2012b; Wang et al. 2013; Béthermin et al. 2013). Lee et al. (2009) studied the UV-light-to-halo-mass ratio at $3 < z < 5$, and found a decrease of this ratio with time at fixed halo mass. Finally, Lin et al. (2012) studied the clustering of $z \sim 2$ galaxies as a function of SFR and sSFR, estimated using UV luminosity corrected for dust extinction, and found clustering increasing with the distance of galaxies from the main sequence (i.e., with sSFR). Finally, several studies using the correlated anisotropies of the cosmic infrared background (CIB), which is the relic emission of the dust emission from all star-forming galaxies across cosmic times, showed that from $z = 0$ to $z = 3$ the bulk of the star formation is hosted by halos of $\sim 10^{12-13}\ M_\odot$ (e.g., Béthermin et al. 2013; Viero et al. 2013; Planck Collaboration XXX 2014). In particular, Béthermin et al. (2013) showed that the strong evolution of populations of star-forming galaxies responsible for the CIB can be modeled assuming a universal efficiency of conversion of accreted cosmological gas into stars as a function of the halo mass peaking around $10^{12.5}\ M_\odot$ and the evolution of the accretion rate at fixed halo mass with time.

In this paper, we study the clustering of individually detected galaxy populations focusing on the $1.5 < z < 2.5$ redshift range when star formation is at its highest level (e.g., Hopkins & Beacom 2006; Le Borgne et al. 2009; Gruppioni et al. 2013; Magnelli et al. 2013; Burgarella et al. 2013; Planck Collaboration XXX 2014) in order to obtain new observational constraints on the link between the star-forming properties of galaxies and the nature of their host halos and environments. Ultimately, our aim is a better understanding of the mechanisms that drive star formation, trigger starbursts and quench galaxies at high redshift.

In Sect. 2, we describe the approach used to build our sample and the estimate of the physical properties of galaxies. In Sect. 3, we detail the method chosen to measure the angular correlation function of our various subsamples and the halo occupation model used to interpret the measurements. Sections 2 and 3 can be skipped by readers not interested in the technical details. In Sects. 4–6, we present and discuss our results on the link between the halo mass and stellar mass, the SFR, and the sSFR, respectively. In Sect. 7, we study the clustering of galaxies depending on their nature, i.e., for the categories main-sequence, starburst, and passive. In Sect. 9, we discuss the consequences of our results on our understanding of galaxy evolution. We conclude in Sect. 10.

In this paper, we assume a WMAP-7 cosmology (Larson et al. 2011) and a Salpeter (1955) initial mass function. We use the virial mass for the definition of the dark matter halo mass.

2. Description of the sample

We built a sample of galaxies at $z \sim 2$ to perform our analysis. We used the BzK selection technique of Daddi et al. (2004), which allows us to efficiently select galaxies around $z = 2$, to split the sample into a star-forming and a passive galaxy population, and to estimate the stellar mass and the SFR using only B-band, z-band, and K-band photometry (see Daddi et al. 2004, 2007 for details). Stellar masses and UV-based SFRs computed in this way have formal errors typically around 0.1–0.2 dex or lower, and within a 0.3 dex scatter are in agreement with those computed from the fit of global spectral energy distributions (see also Rodighiero et al. 2014). We used the same band-merged photometric catalog as in McCracken et al. (2010), selected down to $K_\mathrm{AB} = 23$. However, the UV-derived SFR is not reliable for dust-obscured starbursts (Goldader et al. 2002; Chapman et al. 2005; Daddi et al. 2007). We thus used *Spitzer* and *Herschel*-derived SFR, where available. This ladder of SFR indicators is similar to the one built by Wuyts et al. (2011) and to the one used by Rodighiero et al. (2011).

2.1. COSMOS passive BzK sample

The high-redshift passive galaxies (called hereafter pBzK) are selected using the Daddi et al. (2004) criteria,

$$(z_{AB} - K_{AB}) - (B_{AB} - z_{AB}) \leq -0.2 \text{ and } (z - K)_{AB} > 2.5, \quad (1)$$

where z_{AB}, K_{AB}, and B_{AB} are the magnitude in AB convention of the galaxies in the corresponding bands. To avoid any contamination by low-z interlopers, we also discard galaxies with a photometric redshift (coming from Ilbert et al. 2009) lower than 1.4. The stellar mass is estimated from the K-band photometry and using the $z-K$ color to estimate the mass-to-light ratio following Daddi et al. (2004). The stellar mass is only weakly dependent on the exact redshift of the sources for pBzKs and no correction taking into the photometric redshift is performed.

2.2. COSMOS star-forming BzK sample

The high-redshift, star-forming galaxies (called hereafter sBzK) lie in another part of the BzK diagram (Daddi et al. 2004):

$$(z_{AB} - K_{AB}) - (B_{AB} - z_{AB}) \geq -0.2. \quad (2)$$

We apply the same cut on the photometric redshift ($z > 1.4$) to remove the low-redshift interlopers and use the same method as for pBzK to estimate their stellar mass. The SFR in sBzK is estimated from the B-band photometry, which is corrected for attenuation estimating the UV-slope using the $(B - z)$ color. The SFR estimate is more sensitive to the exact redshift of the source in the interval, and the photometric redshift is used to refine the value of SFR when it is available (95% of the sample). In addition to these criteria, we remove the objects classified as passive in the UVJ diagram (Williams et al. 2009, see also Wuyts et al. 2007) from the sBzK sample.

2.3. Spitzer/MIPS data

In highly obscured, dusty galaxies, only a small fraction of UV light can escape from the interstellar dust clouds hosting young stars and UV-corrected estimates of SFR are not reliable (see discussion in Rodighiero et al. 2011). An estimate of SFR from infrared data is then more reliable. Le Floc'h et al. (2009) built a catalog of 24 μm sources matched with the data from which the BzK sample was built. The total infrared luminosity (L_{IR}, integrated between 8 μm and 1000 μm) is extrapolated from the 24 μm flux density using the Magdis et al. (2012) templates; L_{IR} is then converted into SFR assuming the Kennicutt (1998) conversion factor.

2.4. Herschel/PACS data

At very high 24 μm flux density, the SFR derived from this wavelength is also no longer reliable. On the one hand, AGN contamination becomes significant (e.g., Treister et al. 2006). On the other hand, the ratio between PAH features and the cold dust continuum is lower in starbursting galaxies (Elbaz et al. 2011). For this reason, we use the *Herschel*/PACS catalog of the COSMOS field, which was extracted using the positions of 24 μm sources as a prior. The data comes from the PACS evolutionary probe survey (PEP, Lutz et al. 2011). We fitted the 100 μm and 160 μm PACS flux densities with Magdis et al. (2012) templates in order to derive L_{IR} of each galaxy and assume the same conversion factor between L_{IR} and SFR as for MIPS. The PACS wavelengths have the advantage of being close to the maximum of emission of the dust, and the recovered L_{IR} is few sensitive to the assumed temperature.

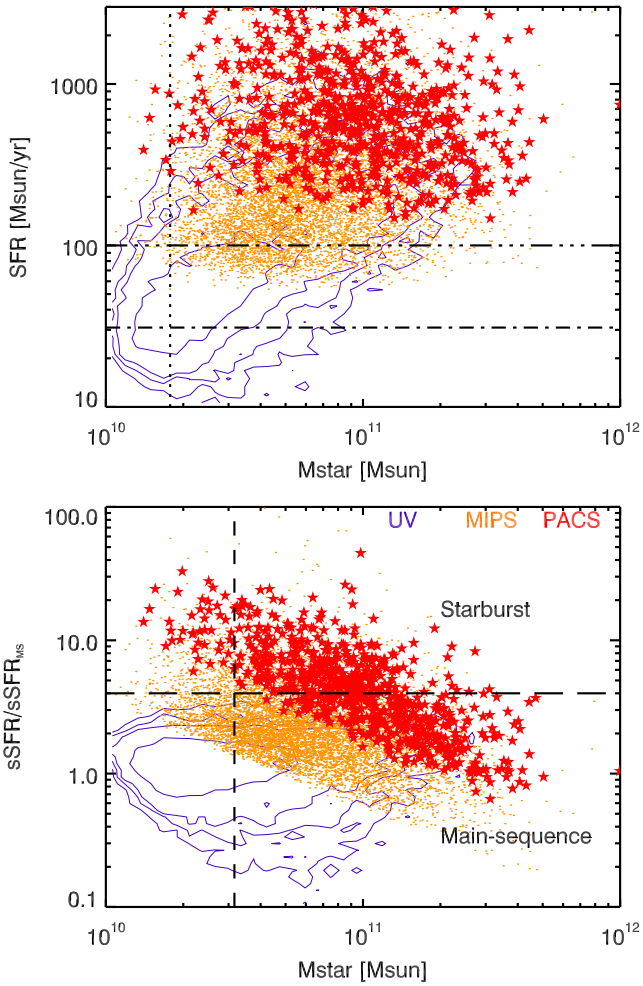

Fig. 1. *Upper panel*: position of our sBzK sources in the *SFR-$M_\star$* plane. The purple contours indicate sources with only UV-detection (i.e., having no mid/far-IR), the orange dots the MIPS-but-not-PACS-detected sources, and the red stars the PACS-detected sources. The vertical dotted line indicates the stellar mass where the sample becomes strongly incomplete. The horizontal dot-dash and three-dot-dash lines represents the completeness limits in SFR of the UV and MIPS subsamples. *Lower panel*: distance between our sBzK sources and the center of the main sequence. The horizontal long-dash line corresponds to the limit between the galaxies classified as main sequence of starburst. The vertical short-dash line shows the mass cut that defines a complete mass-selected sample of starburst galaxies detected by PACS.

2.5. SFR-$M_\star$ diagram

Figure 1 (upper panel) shows the position of our sBzK sources in the classical *SFR-$M_\star$* diagram. The correlation (so-called main sequence) between SFR and $M_\star$ is well probed using UV-derived SFRs. The pBzKs, not shown in the figure, have a low SFR, which is thus very difficult to estimate accurately. They lie well below the main sequence. Because of their high threshold in SFR, the correlation is poorly detected by MIPS and not seen by PACS. This shows why it is so important to use various SFR estimators to accurately probe the full *SFR-$M_\star$* diagram. In this paper, we will present the clustering as a function of various physical parameters ($M_\star$, SFR, sSFR). We have to define various cuts for which our samples are complete.

Below $M_\star = 10^{10.25}\ M_\odot$, the catalog of star-forming BzK becomes incomplete. This is caused by the magnitude limits of the catalog in the B, z, and K bands. Consequently, we will apply these cuts when we study the clustering as a function of stellar mass. For pBzK galaxies, the catalog begins to be incomplete below $M_\star \sim 10^{10.8}\ M_\odot$. We define two samples one of which is highly complete above $10^{11}\ M_\odot$ and we also use a mass bin between $10^{10.5}\ M_\odot$ and $10^{11}\ M_\odot$, albeit somewhat incomplete. Incompleteness is not a substantial problem for clustering studies unless it is correlated with environment. Concerning the selections in SFR, we wanted to define an SFR limit above which the sample is not affected by the mass incompleteness caused by the K-band sensitivity limit. We estimated this SFR cut to be 30 $M_\odot\ \mathrm{yr}^{-1}$ (see Fig. 1 lower panel). MIPS and PACS are only sensitive to $SFR \gtrsim 100\ M_\odot$/yr and $\gtrsim$200 $M_\odot\ \mathrm{yr}^{-1}$, respectively, and are not affected by the incompleteness in mass. For MIPS, we use a sharp cut at 100 $M_\odot\ \mathrm{yr}^{-1}$. For the PACS selected sample, we use the full sample to maximize the statistics because the total number of detections is already small for a clustering study.

The selection in sSFR is more difficult. The same sSFR can be measured in a low mass galaxy hosting a low SFR and in a massive, strongly star-forming galaxies. It is then impossible to define a completeness cut for sSFR. For this reason, we first apply a stellar mass threshold before sorting the galaxies by sSFR. We have chosen a slightly high cut of $10^{10.5}\ M_\odot$, which allows us to detect all the $M_\star > 10^{10.5}\ M_\odot$ starburst galaxies with PACS. The starburst galaxies are defined as being 0.6 dex (a factor of 4) above the main sequence following Rodighiero et al. (2011). There is no clear gap between main-sequence and starburst galaxies in the sSFR distribution. However, Rodighiero et al. (2011) and Sargent et al. (2012) showed evidence for a departure from a log-normal distribution at sSFRs 0.6 dex larger than the center of the main sequence. Modeling studies (Béthermin et al. 2012c; Sargent et al. 2013) suggested that the value of 0.6 dex corresponds to the transition between secularly star-forming galaxies and merger-induced starbursts. Nevertheless, this bimodality of the star-formation modes does not imply a strict separation in the SFR-$M_\star$ diagram. The PACS detection is crucial in our analysis, because starbursts are highly obscured, which makes the UV-derived SFR not very reliable. In addition, they present a PAH deficit (Elbaz et al. 2011), which implies an underestimated SFR derived from 24 μm. Figure 1 (lower panel) shows the distance between the galaxies and the main sequence. We use the same definition as Béthermin et al. (2012a) for the center of the main sequence,

$$sSFR_{MS} = 10^{-10.2}\ \mathrm{yr}^{-1} \times \left(\frac{M_\star}{10^{11}\ M_\odot}\right)^{-0.2} (1+z)^3, \quad (3)$$

where z and $M_\star$ are the best estimate of the photometric redshift (Ilbert et al. 2010) and the stellar mass for each galaxy.

2.6. Redshift and mass distributions

In order to interpret the clustering measurements, we need to know the redshift distribution of our sources. Figure 2 (upper panel) shows the redshift distribution of our various subsamples, i.e., pBzK-selected passive galaxies, sBzK-selected main-sequence galaxies, and sBzK-and-PACS-selected starbursts. For the three categories, the bulk of the source lies between $z = 1.4$ and $z = 2.5$. However, the star-forming galaxies have a tail at $z > 2.5$, while the passive galaxies do not, probably because a small fraction of galaxies (including the massive ones) are already quenched at $z > 2$ (Ilbert et al. 2013; Muzzin et al. 2013). This could also be due to photometric selection effects.

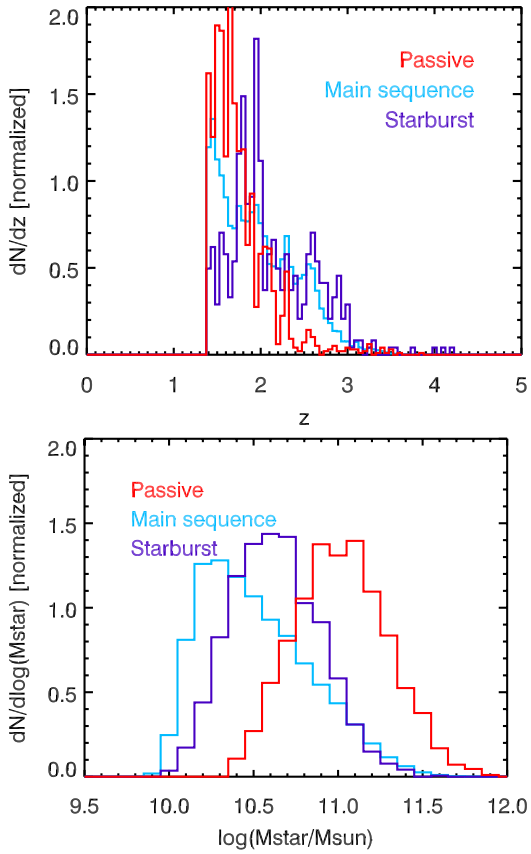

Fig. 2. *Upper panel*: redshift distribution for our samples of passive (red), main-sequence (light blue), and starburst (purple) galaxies. *Lower panel*: stellar-mass distributions.

The distribution of stellar mass of the three subsamples is shown Fig. 2 (lower panel). The rapid decrease in the number of main-sequence and starburst galaxies below $\sim 10^{10}\ M_\odot$ is caused by the magnitude limit of the survey in K band. The small offset ($\sim$0.2 dex) between the position of this break for these two populations is probably caused by a larger attenuation in the starbursts, and the FIR depth required to select starbursts. The mass distribution of passive galaxies peaks around $10^{11}\ M_\odot$ in agreement with Ilbert et al. (2013) and Muzzin et al. (2013), but we caution that in our sample this happens fairly close to the completeness limit.

3. Clustering measurements and modeling

3.1. Measurements of the correlation function

One of the simplest estimators of the clustering of galaxies is the autocorrelation function (ACF), often written as $w(\theta)$. This function is the excess probability over Poisson of finding a neighbor at an angular distance θ from another source. This function can be measured using the Landy & Szalay (1993) estimator,

$$w(\theta) = \frac{1}{RR} \times \left(\frac{N_\mathrm{r}(N_\mathrm{r}-1)}{N_\mathrm{d}(N_\mathrm{d}-1)} DD - 2\frac{N_\mathrm{r}}{N_\mathrm{d}} DR + RR\right), \quad (4)$$

where N_r and N_d are the number of objects in the random and galaxy catalogs, respectively; DD, DR, and RR are the number of pairs, separated by an angle between $\theta - \mathrm{d}\theta/2$ and $\theta + \mathrm{d}\theta/2$,

in the galaxy catalog, between the random and the galaxy catalogs, and in the random catalog, respectively. The random catalog contains 100 000 objects to minimize the Poisson errors coming from the statistical fluctuations of DR and RR. It was drawn in a uniform way in the field excluding the masked regions. The angular, cross-correlation function χ_{AB} between two population (A and B) is computed with

$$\chi_{AB}(\theta) = \frac{1}{R_A R_B} \times \left(\frac{N_{rA} N_{rB}}{N_A N_B} D_A D_B - \frac{N_{rB}}{N_A} D_A R_B - \frac{N_{rA}}{N_B} D_B R_A \right) + R_A R_B, \quad (5)$$

where N_A is the number of galaxies in the population A (N_B in population B), N_{rA} the number of sources in the random catalog A (N_{rB} in the random catalog B), $D_A D_B$ the number of pairs between the population A and B, and $D_A R_B$ ($D_B R_A$, respectively) between population A (B, respectively) and the random catalog B (A, respectively). These computations were performed with the ATHENA code[1]. As Coupon et al. (2012) did, we estimated the uncertainties using a jackknife method, splitting the field into 20 subfields. This technique also allows us to estimate the covariance matrix between the angular bins, which is used to fit the models (see Sect. 3.3.1).

The Landy & Szalay (1993) estimator is biased by an offset called integral constraint C ($w_{mes} = w - C$, where w_{mes} is the measured ACF and w the intrinsic one),

$$C = \frac{1}{\Omega^2_{field}} \int_{field} \int_{field} w(\theta) \, d\Omega_1 \, d\Omega_2, \quad (6)$$

where Ω_{field} is the size of the field and θ is the distance between two points drawn randomly in the field. The double integral takes into account all the possible random pairs in the field. In practice, we pre-computed the angles between 10^6 pairs randomly taken in the field. Before comparing a model to the data, we compute C taking the mean $w(\theta)$ value provided by the model for these angles. The method is exactly the same for the cross-correlation function.

3.2. Estimation of the correlation length

A simple way to interpret the clustering of a population is to determine its correlation length r_0. We will estimate this value for our various samples to allow an easy comparison with the literature. The 3D angular correlation function is assumed to be described by the power law

$$\epsilon(r) = \left(\frac{r}{r_0} \right)^{-\gamma}, \quad (7)$$

where r is the comoving distance between the two points and r_0 the characteristic correlation length; γ is fixed at the standard value of 1.8.

In the flat-sky approximation of Limber (1953), the corresponding angular correlation function is then (Peebles 1980),

$$w(\theta) = \theta^{1-\gamma} r_0^{\gamma} \frac{\Gamma(\gamma/2)}{\Gamma(1/2)\Gamma((\gamma-1)/2)} \frac{c}{H_0} \frac{\left(\int \frac{dN}{dz} dz \right)^2}{\int E(z) D_C^{1-\gamma} \left(\frac{dN}{dz} \right)^2 dz}, \quad (8)$$

where H_0 is the Hubble constant, $E(z) = \sqrt{(1+z)^3 \Omega_m + \Omega_\Lambda}$, D_C the comoving distance corresponding to a redshift z, and dN/dz the redshift distribution of the sample. We used the redshift distribution of passive, main-sequence, or starburst galaxies presented in Sect. 2.6. We checked that using the redshift distribution of the full parent population, i.e. all passive, all main-sequence, or starburst galaxies depending on the case, or only the distribution of mass- or SFR-selected subsamples, does not bias the results significantly ($>1\sigma$). We thus used the redshift distributions of the full parent samples in our analysis to avoid any problems caused by low statistics for the small subsamples. On small scales, the correlation function is not exactly a power law and exhibits an excess caused by the correlation of galaxies inside the same dark matter halo. This is especially relevant for passive galaxies as discussed in Sect. 7. We thus fit only scales larger than 30 arcsec to determine r_0.

[1] http://www2.iap.fr/users/kilbinge/athena/

3.3. Determination of the mean mass of the host halos

We also used a standard halo occupation distribution (HOD) model (see Cooray & Sheth 2002, for review), described in Sect. 3.3.1, and a method based on the large-scale clustering, explained in Sect. 3.3.2, to estimate the typical mass of halos hosting the various subsamples we studied in this paper. The consistency of these methods and the validity of the assumptions on which they are based is discussed in Sect. 3.3.3.

3.3.1. With a halo occupation model

The HOD models interpret the clustering of galaxies using a parametric description of how galaxies occupy halos. We follow the same approach as Coupon et al. (2012), based on the Zheng et al. (2005) formalism. This formalism was initially built to study samples selected using a mass threshold. In Sect. 3.3.3, we discuss the pertinence of this formalism for our samples.

In the HOD formalism, the total number of galaxies in a halo of mass M_h is described by

$$N(M_h) = N_c \times (1 + N_s(M_h)), \quad (9)$$

where N_c is the number of central galaxies and N_s the number of satellites. These two quantities are assumed to vary only with halo mass. Classically, N_c is parametrized by

$$N_c(M_h) = \frac{1}{2} \left(1 + \mathrm{erf} \left(\frac{\log(M_h) - \log(M_{min})}{\sigma_{\log(M)}} \right) \right), \quad (10)$$

where M_{min} is the minimal mass of halos hosting central galaxies and $\sigma_{\log(M)}$ is the dispersion around this threshold. Another parametrization is used for N_s,

$$N_s(M_h) = \left(\frac{M_h - M_0}{M_1} \right)^{\alpha}, \quad (11)$$

where M_1, M_0, and α are free parameters of the HOD models.

The autocorrelation function can be derived using the classical formula

$$w(\theta) = \frac{\int_z \left(\frac{dN}{dz} \right)^2 \int_k \frac{k}{2\pi} P_{gg}(k,z) J_0(k D_c \theta) \, dz \, dk}{\left(\int_z \frac{dN}{dz} \, dz \right)^2}, \quad (12)$$

where J_0 is the zero-th order Bessel function and dN/dz the redshift distribution of the galaxy sample (see Sect. 2.6); P_{gg} is the sum of two terms corresponding to the clustering of galaxies in two different halos and inside the same halo:

$$P_{gg}(k,z) = P^{2h}_{gg}(k,z) + P^{1h}_{gg}(k,z). \quad (13)$$

Following the standard method, we compute the two-halo term with

$$P_{\rm gg}^{\rm 2h}(k,z)=\left[\int_{M_{\rm h}}\frac{{\rm d}^2N}{{\rm dlog}(M_{\rm h}){\rm d}V}b(M_{\rm h})\frac{N_{\rm c}+N_{\rm s}}{\bar{n}_{\rm gal}}{\rm d}M_{\rm h}\right]^2 P_{\rm lin}(k,z), \quad (14)$$

where ${\rm d}^2N/{\rm d}\log(M_{\rm h}){\rm d}V$ is the halo mass function and $\bar{n}_{\rm gal} = \int\frac{{\rm d}^2N}{{\rm d}\log(M_{\rm h}){\rm d}V}(N_{\rm c}+N_{\rm s}){\rm d}M_{\rm h}$ the mean number of galaxies. The one-halo term is computed with

$$P_{\rm gg}^{\rm 1h}(k,z) = \int_{M_{\rm h}}\frac{{\rm d}^2N}{{\rm d}\log(M_{\rm h}){\rm d}V}\frac{2N_{\rm c}N_{\rm s}+N_{\rm s}^2}{\bar{n}_{\rm gal}^2}u^2(k,M_{\rm h},z){\rm d}M_{\rm h}, \quad (15)$$

where u is the Fourier transform of the NFW halo density profile (Navarro et al. 1997).

We fitted the measured correlation functions using a Markov Chain Monte Carlo (MCMC) Metropolis Hastings procedure. We used the same flat priors as in Coupon et al. (2012). We used only the angular distance larger than 3″ to avoid any bias caused by an incorrect deblending. We did not use the number of galaxies ($\bar{n}_{\rm gal}$) as a constraint, since we considered only subsamples of the full galaxy population at $z \sim 2$ (see discussion in Sect. 3.3.3). Taking this constraint into account (as in Wetzel et al. 2013) would request some extra hypotheses and a large number of extra free parameters, which are too hard to constrain at high redshift. The results could also be potentially biased by the choices of parametric forms used to describe the halo occupation by the various populations. Nevertheless, we compared our results a posteriori with the measurements based on the abundance matching technique and found a good agreement (see Sect. 4). Strong degeneracies exists between the various HOD parameters. For this reason, we consider in this paper only the mean host halo mass mass ($\langle M_{\rm h}\rangle$) defined as

$$\langle M_{\rm h}\rangle = \int M_{\rm h}\frac{{\rm d}^2N}{{\rm d}\log(M_{\rm h}){\rm d}V}\frac{N_{\rm c}+N_{\rm s}}{\bar{n}_{\rm gal}}{\rm d}M_{\rm h}. \quad (16)$$

At each step of the MCMC, we save this quantity, which is derived from the five HOD parameters, in order to build the confidence region. However, the probability distribution of $\langle M_{\rm h}\rangle$ is not flat if we pick random vectors in the 5D HOD parameter space following the priors. For this reason, we divided the probability distribution found by MCMC by the one recovered drawing 100 000 random sets of HOD parameters.

3.3.2. Using the large-scale clustering

Subsamples requiring a PACS selection contain in general a small number of objects and their clustering cannot be estimated accurately at small scale because of the size of the PACS PSF (12″ at 160 μm). In these particular cases, we used only the large scales (>30″, corresponding to 0.8 Mpc comoving at $z = 2$), where the one-halo term can be neglected, to constraint the effective linear bias of these populations of galaxies. This bias is then converted into halo mass assuming the $M_{\rm h}$-bias relation of Tinker et al. (2008). If we assume a constant bias ($b_{\rm pop}$) over the full redshift range, the two-halo term is then given by

$$w(\theta) = b_{\rm pop}^2 \times w_{\rm DM,pop} \quad (17)$$

with

$$w_{\rm DM,pop} = \frac{\int_z\left(\frac{{\rm d}N_{\rm pop}}{{\rm d}z}\right)^2\int_k\frac{k}{2\pi}P_{\rm lin}(k,z)J_0(kD_{\rm c}\theta)\,{\rm d}z\,{\rm d}k}{\left(\int_z\frac{{\rm d}N_{\rm pop}}{{\rm d}z}\,{\rm d}z\right)^2}. \quad (18)$$

We first computed $w_{\rm DM,pop}$, then fitted $b_{\rm pop}$ to the measured ACF, and finally determined the halo mass corresponding to this bias.

3.3.3. Validity and consistency of the two approaches

The two methods presented here have various pros and cons. The HOD approach treats all the scales at the same time and avoids artificially cutting the data at a scale where the one-halo term is supposed to be negligible. The method based on the large-scale clustering is simpler, but could be biased by residuals of the one-halo term. The error bars are similar for both methods, even if the HOD method is based on more data points. This is caused by the degeneracy between the one-halo and the two-halo terms, since the large-scale clustering method assumes no one-halo term. We used HOD modeling when it is possible to constrain sufficiently well the one-halo, i.e., all our subsamples except the PACS-selected one.

The classical HOD formalism assumes implicitly that the sample is selected by applying a threshold and that the quantity used to define this threshold is correlated with the halo mass. This assumption is not exact for our subsamples (e.g., SFR-selected sample, sSFR-selected samples, selection by interval and not threshold). We could modify the HOD parametrization, but we cannot know which parametric representation is the most correct before studying the clustering of these populations. However, several problems can happen with the classical HOD applied to our subsamples. First, our subsamples occupy only a fraction of the halos and the number of central galaxies never reaches 1. This is not a problem for the HOD modeling if we consider only the clustering because w(θ) stays the same if we multiply $N_{\rm c}$ and $N_{\rm s}$ by the same factor f. This would be a problem if we had used the $\bar{n}_{\rm gal}$ constraint. In addition, at large halo mass, for star-forming samples, the mean number of central galaxies could tend to zero because central galaxies of massive halos are often passive. This is compensated for artificially by the HOD model, which underestimates the number of satellites to compensate an excess of centrals, since they play a symmetric role in the equations. However, this would be a problem if we had attempted to measure the fraction of satellites, which is studied in another paper using samples selected with a mass threshold (Wolk et al., in prep.).

We checked the consistency between the two methods comparing the recovered mean halo mass for all the samples used in this paper and for which both methods can be applied. The result is shown Fig. 3. We fitted the data with a linear relation and found $\log(M_{\rm h}^{\rm HOD}/10^{12}\ M_\odot) = (0.91 \pm 0.23) \times \log(M_{\rm h}^{\rm LSB}/10^{12}\ M_\odot) + (0.07 \pm 0.12)$ with a reduced χ^2 of 0.59. The slope is thus compatible with unity and the offset is small. If we force the slope at 1, the offset is only 0.02 dex. There is thus a very good consistency between these two methods, which suffer different systematics. This indicates that at the level of precision reached in this paper, the assumptions we made are reasonable. This is not surprising because the mean halo mass found by the HOD is strongly related to the large-scale clustering, which is driven by the effective bias of the halos hosting the population (see Eq. (14)).

4. Relation between halo and stellar mass

4.1. Results

We measured the ACF for various subsamples of sBzK and pBzK sorted by stellar mass. For the star-forming galaxies, we used the following bins: $10.25 < \log(M_\star/M_\odot) < 10.50$, $10.50 < \log(M_\star/M_\odot) < 10.75$, $10.75 < \log(M_\star/M_\odot) < 11.00$, $11.00 < \log(M_\star/M_\odot) < 11.25$, and $11.25 < \log(M_\star/M_\odot) < 11.50$. For the passive galaxies, their number density is smaller and we thus

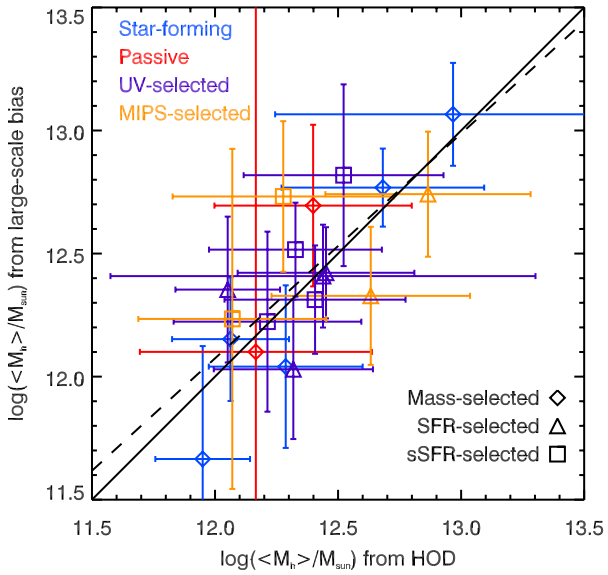

Fig. 3. Comparison between the mean host-halo mass derived from the determination of the large-scale (>1′) bias (see Sect. 3.3.2) and from HOD modeling (see Sect. 3.3.2). The data points represent all the samples for which the HOD model can be applied, i.e., the non-PACS-selected ones. The solid line is the one-to-one relation and the dashed line the best, linear fit of the data.

used larger mass bins in order to have a sufficient signal: $10.5 < \log(M_\star/M_\odot) < 11.0$ and $11.0 < \log(M_\star/M_\odot) < 11.5$. The results are shown in Fig. 4. The results are well fitted by the power-law model at $\theta > 30''$ (0.0083°, vertical dotted line). On smaller scale, we detect some excess due to the one-halo clustering, especially for passive subsamples. The HOD model is very flexible and thus nicely fits the data. Figure 5 shows the resulting correlation length (r_0, upper panel) and mean halo mass (lower panel). These two quantities increase with the stellar mass.

We compared our results on the correlation length with the results of Lin et al. (2012) (Fig. 5, upper panel). Our results are systematically lower than theirs. However, their analysis was based on the GOODS-N field, which is much smaller than COSMOS (150 versus 7200 arcmin^2). They fitted scales between 3.6″ and 6′, while we focused on the 30″ to 12′ range. Consequently, their analysis is more sensitive to the intra-halo clustering, and our analysis is more sensitive to the large-scale clustering. The fact that their values are higher than ours is consistent with the excess of clustering at small scales compared to our power-law fit of the large scales. Their error bars are just slightly larger than ours despite a field that is 13 times smaller. This is essentially caused by their use of the scales below 30″.

4.2. Interpretation

We found similar clustering lengths to be similar for passive and star-forming galaxies at 1σ. This is in agreement with the results of Wetzel et al. (2013) at $z < 1$, who found that the mean halo mass at fixed stellar mass is similar for both populations. This also justifies a posteriori the hypothesis of the same $M_\star - M_\mathrm{h}$ relation for both populations in the model of Béthermin et al. (2013) linking dark matter halos and infrared galaxy populations. The correlation length is compatible with the prediction of the Lagos et al. (2011) model, contrary to what was claimed by Lin et al. (2012) because they used a very small scale signal for which the one-halo term could be in excess compared to

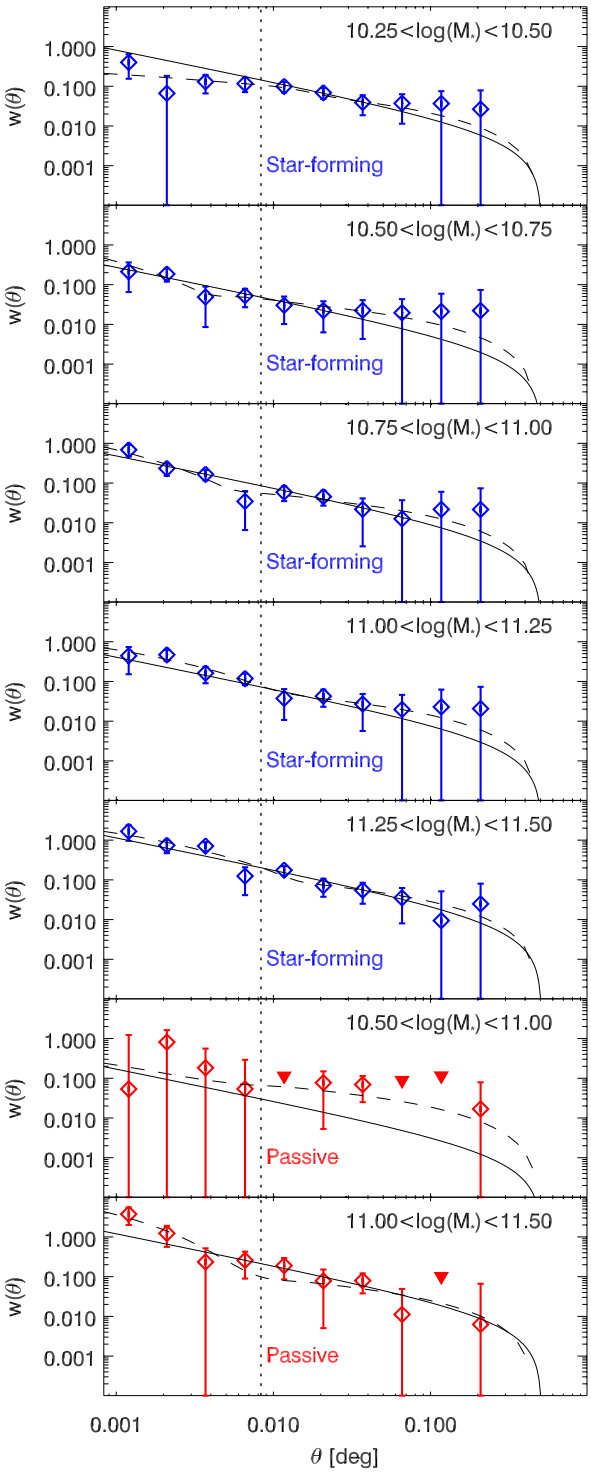

Fig. 4. Autocorrelation function of our various subsamples of star-forming (blue) and passive (red) galaxies selected by stellar mass. The solid line is the best fit for the power-law model (fitted only above the 30″ limit represented by the vertical dotted line) and the dashed line the best-fit HOD model. The triangles represent the 3σ upper limits for data points with negative central values, which cannot be represented in a logarithmic plot.

the power-law approximation and because of possible problems of deblending. We also compared our $M_\star - M_\mathrm{h}$ relation with the results of estimates based on abundance matching (Behroozi et al. 2010; Moster et al. 2010; Béthermin et al. 2012b) finding

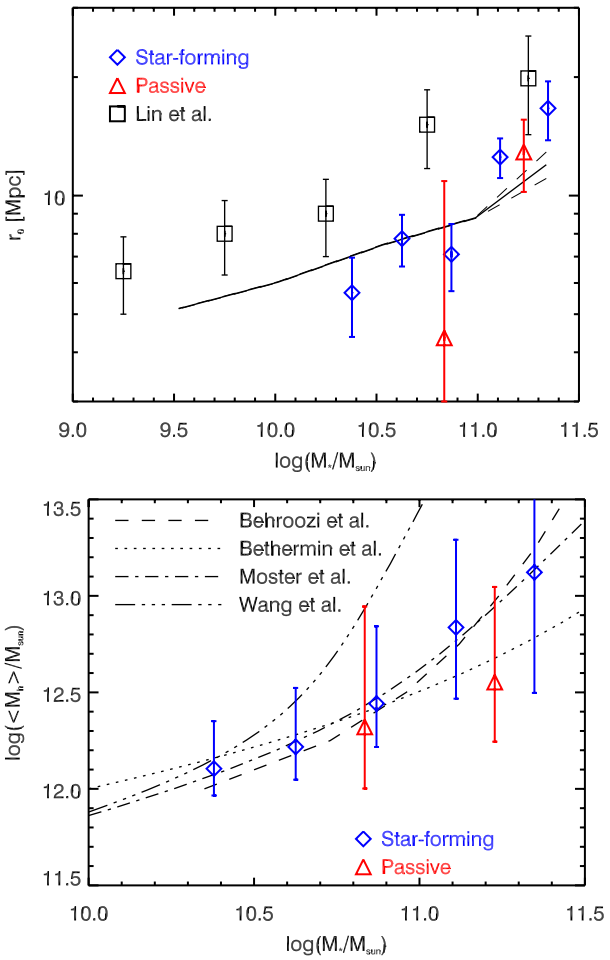

Fig. 5. *Upper panel*: correlation length of star-forming (blue diamonds) and passive (red triangles) galaxies as a function of their stellar mass. For comparison, the data from Lin et al. (2012) are plotted as black squares. The predictions of the Lagos et al. (2011) model is represented with a black solid line, and the 1σ confidence region with black dashed line. *Lower panel*: mean mass of halos, which host star-forming (blue diamonds) and passive (red triangles) galaxies as a function of their mass. For comparison, we over-plotted the estimates performed by abundance matching by Behroozi et al. (2010; dashed line), Béthermin et al. (2012b; dotted line), Moster et al. (2010; dot-dash line), and Wang et al. (2013).

general agreement, except for the case of Wang et al. (2013) who predicts larger halo mass for $M_\star > 10^{11}\ M_\odot$ galaxies. However, they predicted mean stellar mass at fixed halo mass, while we measured the inverse. This agreement suggests that the basic hypothesis of the abundance matching, that there is a monotonic relation between halo and stellar mass, is still valid at $z = 2$.

5. Relation between halo mass and star formation rate

5.1. Results

We then studied the mean halo mass as a function of the SFR. As explained in Sect. 2, we used by order of priority PACS, MIPS, and UV-derived SFR. We notice that some non-PACS-detected objects have UV-derived and MIPS-derived SFR well above the PACS limit. These sources have an incorrect dust correction for UV SFR and/or an AGN contamination in their 24 μm SFR. These sources contaminate the bright flux samples and we excluded them from samples where the SFR limit is sufficiently high that the catalog should be complete using only MIPS and/or PACS detected sources, as these objects have likely been overcorrected for attenuation. Consequently, we consider several samples: one containing all the sources and using the best available estimator (called hereafter the UV sample), one using only PACS or MIPS by order of priority (called the MIPS sample), and a sample with PACS only. We checked the consistency between the results in the overlapping regions.

We used the following bins with similar sizes in logarithmic unit: 31 $M_\odot\,\mathrm{yr}^{-1} < SFR < 56\ M_\odot\,\mathrm{yr}^{-1}$, 56 $M_\odot\,\mathrm{yr}^{-1} < SFR < 100\,M_\odot\,\mathrm{yr}^{-1}$, 100 $M_\odot\,\mathrm{yr}^{-1} < SFR < 177\ M_\odot\,\mathrm{yr}^{-1}$, and 177 $M_\odot\,\mathrm{yr}^{-1} < SFR < 316\ M_\odot\,\mathrm{yr}^{-1}$ for the UV sample; and 100 $M_\odot\,\mathrm{yr}^{-1} < SFR < 213\ M_\odot\,\mathrm{yr}^{-1}$ and 213 $M_\odot\,\mathrm{yr}^{-1} < SFR < 457\ M_\odot\,\mathrm{yr}^{-1}$ for the MIPS sample. The PACS sample is too small to be split into several subsamples. We consequently used the full sample. This selection is roughly similar to a $SFR > 200\ M_\odot\,\mathrm{yr}^{-1}$ selection. Finally, we built a subsample of PACS-detected main-sequence galaxies, removing the starbursts from the previous sample. The results are presented in Fig. 6. The data of UV- and MIPS-selected samples are well fitted by both the power-law and the HOD models. For the PACS-selected samples, we used a power-law fit. The PACS full sample has few objects and is weakly clustered. Consequently the signal is poorly detected and there are a similar number of positive and negative measurements (three each). However, the negative points are all close to zero and there is a 2σ positive outlier. We thus obtain a mean positive signal at $\sim 1\sigma$ by fitting these six points. The clustering signal of PACS-detected, main-sequence galaxies is detected with higher signal-to-noise ratio, apparently because the lower number of objects is largely compensated for by a much stronger clustering.

Figure 7 presents the correlation length (upper panel) and the mean halo mass (lower panel) as a function of the SFR. Our results disagree with Lin et al. (2012), (they looked at smaller scales as explained in the previous section). At $SFR < 200\ M_\odot\,\mathrm{yr}^{-1}$, we see evidence of a rise in the correlation length and the mean host halo mass with SFR at $\sim 2\sigma$ ($r_0 \propto SFR^{0.32\pm0.21}$ and $M_\mathrm{h} \propto SFR^{1.2\pm0.6}$). At higher SFR, the data are compatible with a plateau ($r_0 \propto SFR^{0.0\pm0.2}$ and $M_\mathrm{h} \propto SFR^{0.2\pm0.6}$ for $SFR > 100\ M_\odot\,\mathrm{yr}^{-1}$ data points). This flattening of the $SFR - M_\mathrm{h}$ relation is thus significant at only 1.7σ, and future analyses on larger samples will be necessary to confirm wether this trend is real or just a statistical fluctuation. However, if we remove the starbursts from the PACS sample, r_0 and $\langle M_\mathrm{h}\rangle$ keep rising above $200\ M_\odot\,\mathrm{yr}^{-1}$. We also compared our results with the model of Lagos et al. (2011), which agrees with the data around $SFR \sim 50\ M_\odot$/yr. Unfortunately, this model predicts very few objects with $SFR > 50\ M_\odot$/yr, and no prediction can be done above this cut.

5.2. Interpretation

The evolution of the SFR-$\langle M_\mathrm{h}\rangle$ is more difficult to interpret than the $M_\star$-$\langle M_\mathrm{h}\rangle$. The increasing mean halo mass with SFR below $200\ M_\odot\,\mathrm{yr}^{-1}$ is a combined consequence of the monotonic relation between stellar and halo mass, and the correlation between SFR and $M_\star$ for main-sequence galaxies. At higher SFR, the data seem to indicate at 2σ the presence of a plateau, or at least a flattening. This regime of SFR is dominated by starburst galaxies, which are above the main-sequence as discussed in Sargent et al. (2012). In their framework (2SFM), the bright-end of the SFR function is thus caused by galaxies close to the break of the mass function, with a strong excess of *sSFR* (>0.6 dex).

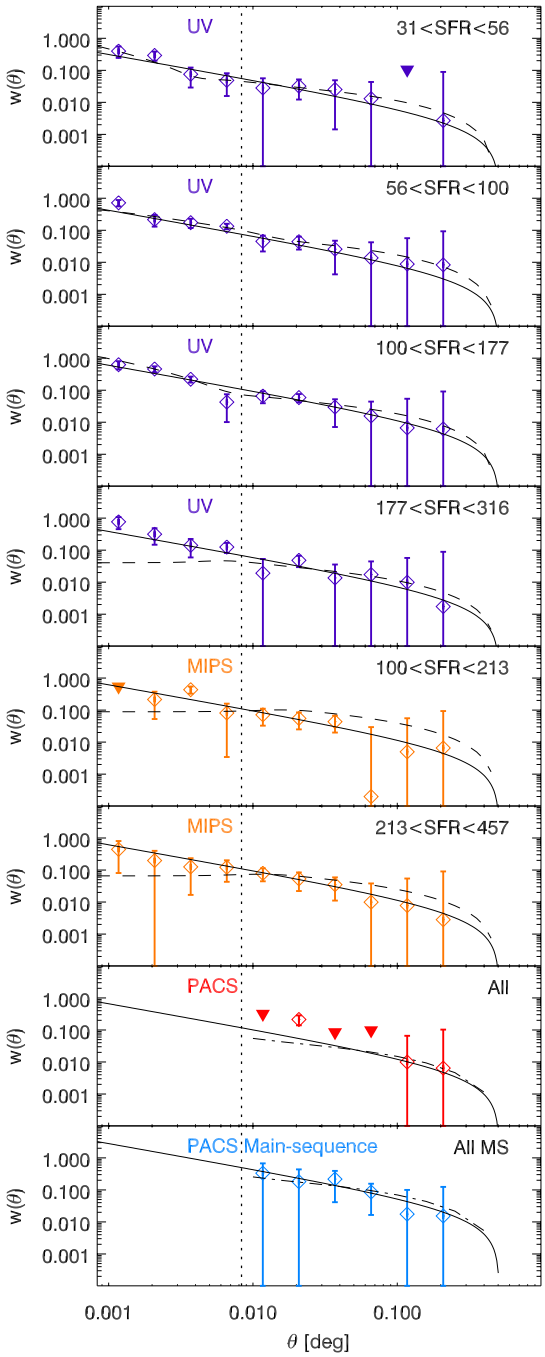

Fig. 6. Autocorrelation function of our various subsamples as a function of their SFR measured with different tracers (purple for UV, orange for MIPS 24 μm, red for PACS, and cyan for PACS main-sequence galaxies). The solid line is the best fit for the power-law model (fitted only above the 30″ limit represented by the vertical dotted line) and the dashed line the best-fit HOD model. The dot-dashed line shows the fit of the linear model for the PACS samples. The triangles represent the 3σ upper limits for data points with negative central values.

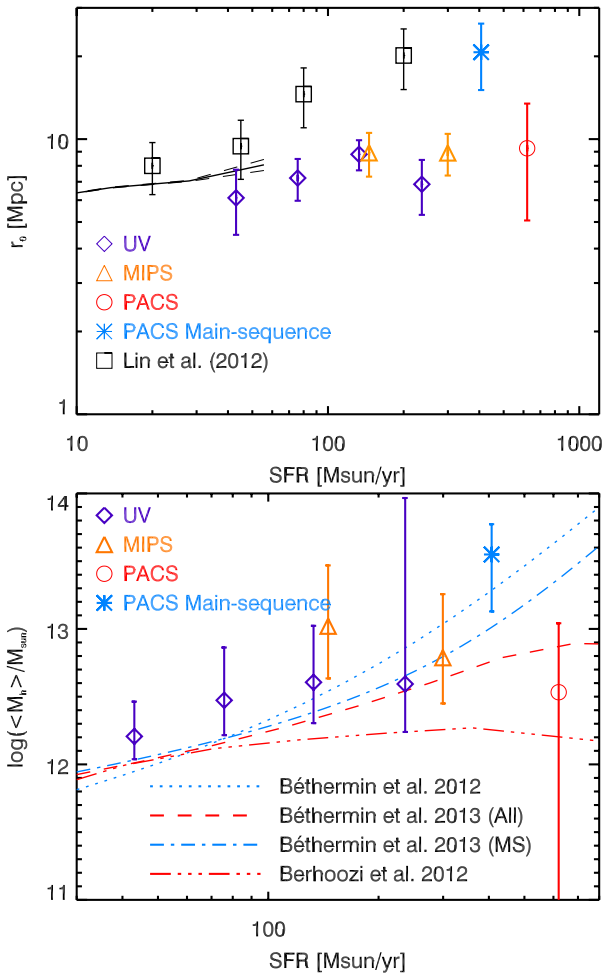

Fig. 7. *Upper panel*: correlation length of our various subsamples as a function of their SFR measured at various wavelengths. For comparison, the data from Lin et al. (2012) are plotted with black squares. The predictions of the Lagos et al. (2011) model is represented with a black solid line, and the 1σ confidence region with a black dashed line. *Lower panel*: mean mass of halos that host galaxies as a function of their star formation rate. We overplotted the estimate performed by abundance matching for main-sequence galaxies of Béthermin et al. (2012b; dotted line), the predictions of the Béthermin et al. (2013) model for main-sequence (dot-dash line) and all (dashed line) galaxies, and the prediction for all galaxies of the Behroozi et al. (2013) model (three-dot-dash line).

Above 300 $M_\odot\,\mathrm{yr}^{-1}$, the mean stellar mass no longer increases with SFR. This flattening thus suggests that the stellar mass is better correlated to the halo mass than the SFR. The flattening of the $SFR - \langle M_\mathrm{h} \rangle$ relation at the same SFR is also an interesting hint (1.7σ) of a modification of the star-formation regime at $z \sim 2$ around 200 $M_\odot\,\mathrm{yr}^{-1}$.

We compared our results with the predictions of various empirical models (Fig. 7, lower panel). Béthermin et al. (2012b) started from the observed stellar mass function and infrared luminosity function (after removing the starburst galaxies) and derived the link between SFR and halo mass for main-sequence galaxies using an abundance matching technique. Their results agree well with our measurements at $SFR < 200\ M_\odot\,\mathrm{yr}^{-1}$, where we can assume that most of the galaxies lies on the main sequence (Sargent et al. 2012), and with the data point corresponding to PACS-detected main-sequence galaxies. Béthermin et al. (2013) proposed an extended version of this approach taking into account both the starburst and quiescent galaxies. Below 200 $M_\odot\,\mathrm{yr}^{-1}$, the model provides very similar predictions, if we consider all galaxies or only objects on the main sequence. At higher SFR, the relation for main-sequence galaxies is steeper, while the relation for all galaxies flattens. These results agree with the difference of clustering we observe between

the all-PACS-detection and PACS-detected, main-sequence samples. Finally, we compared our results with the predictions from the Behroozi et al. (2013) model, which was calibrated from the evolution of the stellar mass function. This model predicts a much flatter relation and tends to be slightly lower than the data around 100 $M_\odot$ yr^{-1} (2σ below the MIPS point at 150 $M_\odot$ yr^{-1}).

We thus found a typical host halo mass for PACS-detected galaxies of $\sim 10^{12.5}$ $M_\odot$. This is about one order of magnitude lower than the measurements of Magliocchetti et al. (2011), who found $10^{13.7^{+0.3}_{-0.4}}$ $M_\odot$ for PACS-detected sources at $z > 1.7$. This could be explained by the fact they used the Mo & White (1996) formalism including scales where the intra-halo clustering is dominant and the potential important cosmic variance caused by the small size of the GOODS-S field. If this is confirmed by future observations on larger fields (e.g., CCAT), this lower value found by our study implies a much lower gap between the typical host halos of local and $z \sim 2$ star-forming galaxies mentioned by Magliocchetti et al. (2014), who consider it to be a clue that high-redshift star-forming galaxies do not have the same nature as the local one. Our new results are more in agreement with models based on the idea of a main sequence evolving continuously with redshift, and with a halo mass where star formation is the most efficient evolving very slowly with redshift (Béthermin et al. 2012b; Behroozi et al. 2013; Béthermin et al. 2013).

6. Relation between halo mass and specific star formation rate

6.1. Results

Finally, we looked at the clustering as a function of sSFR for a stellar mass limited sample ($\log(M_\star/M_\odot) > 10.5$). The mean $\log(M_\star/M_\odot)$ is quite similar for all the subsamples and is between 11.75 and 11.88. We used the same UV, MIPS, and PACS samples as in the previous section. The lower sSFR for which the sample is complete is obtained dividing the SFR cut by the mass cut. The sSFR bins are $sSFR < 0.60\,\mathrm{Gyr}^{-1}$, $0.60\ \mathrm{Gyr}^{-1} < sSFR < 1.50\ \mathrm{Gyr}^{-1}$, $1.50\ \mathrm{Gyr}^{-1} < sSFR < 3.76\ \mathrm{Gyr}^{-1}$, and $3.76\ \mathrm{Gyr}^{-1} < sSFR < 5.96\ \mathrm{Gyr}^{-1}$ for the UV samples; $3\ \mathrm{Gyr}^{-1} < sSFR < 6\ \mathrm{Gyr}^{-1}$, $6\ \mathrm{Gyr}^{-1} < sSFR < 12\ \mathrm{Gyr}^{-1}$, and $12\ \mathrm{Gyr}^{-1} < sSFR < 24\ \mathrm{Gyr}^{-1}$ for the MIPS sample; and $sSFR > 10\ \mathrm{Gyr}^{-1}$ for the PACS sample. Figure 8 shows the clustering measurements and their fit by the power-law and the HOD models. Our results exhibit a flat relation between sSFR and the correlation length (Fig. 9) in disagreement with Lin et al. (2012), who found a minimum of r_0 for an sSFR corresponding to the center of the main sequence (see the arrow in the plot corresponding to the position of the center of the main sequence at $z = 2$ for $\log(M_\star/M_\odot) = 10.5$).

6.2. Interpretation

Lin et al. (2012) interpreted the excess by the fact that galaxies below and above the main sequence are associated with dense environments. This disagrees with the flat relation we find between the sSFR and both r_0 and M_h. Our study is based on more massive galaxies ($\log(M_\star/M_\odot) > 10.5$ versus $\log(M_\star/M_\odot) > 9.5$), but they checked in their analysis that this trend is not mass dependent. The main cause of the difference is probably the fact that we focused on larger scales (>30″), which are essentially associated with the linear clustering and the host halo mass, while they focused on smaller scales (down to 0.3″), which are very sensitive to one-halo clustering and close environmental effects. To check this hypothesis, we made a new fit using scales down to 2″ and found a larger r_0 than with the small scale. This new value is halfway between the Lin et al. (2012) data and our measurements at large scales. At smaller scales and especially below 1″, the autocorrelation function in COSMOS is

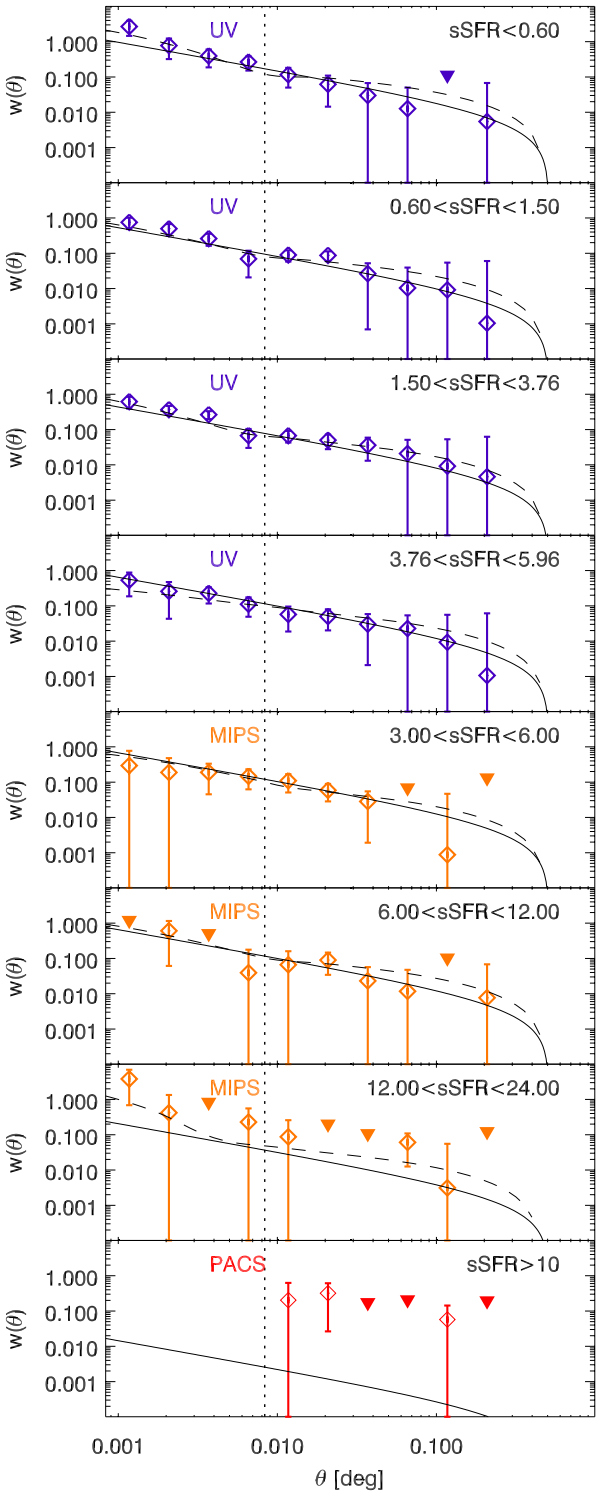

Fig. 8. Autocorrelation function of our various subsamples as a function of their sSFR measured at various wavelengths (purple for UV, orange for MIPS 24 μm, and red for PACS). The solid line is the best fit for the power-law model (fitted only above the 30″ limit represented by the vertical dotted line) and the dashed line the best-fit HOD model. The triangles represent the 3σ upper limits for data points with negative central values.

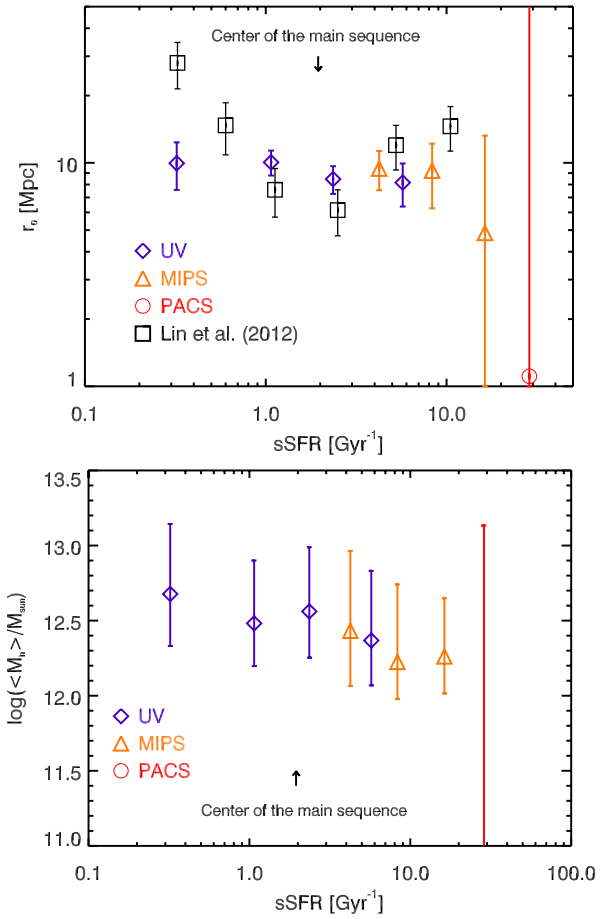

Fig. 9. *Upper panel*: correlation length of our various subsamples of $\log(M_\star/M_\odot) > 10.5$ galaxies as a function of their sSFR measured at various wavelengths (purple for UV, orange for MIPS 24 μm, and red for PACS). For comparison, the data from Lin et al. (2012) are plotted with black squares. *Lower panel*: mean mass of halos, which host $\log(M_\star/M_\odot) > 10.5$ galaxies as a function of their sSFR.

much smaller than their measurements. This could also be partially caused by systematic effects caused by the different deblending methods. In the next section, we will study in detail the clustering of galaxies split into three populations (passive, main sequence, and starburst) to refine our interpretation of their clustering and understand better this tension.

7. Different clustering properties for main-sequence, starburst and passive populations

7.1. Auto- and cross-correlation functions

In this section, we study potential differences between the clustering of galaxies depending of their mode of star formation. We split our $z \sim 2$ sample in three mass-selected ($M_\star > 10^{10.5}\ M_\odot$) subsamples: the passive (the pBzK sample), main-sequence (sBzk with $\Delta\,sSFR_{\rm MS} < 4$), and starburst (PACS-detected sBzK with $\Delta\,sSFR_{\rm MS} > 4$) galaxies. The main-sequence and starburst samples have very close mean stellar masses (10.84 versus 10.81). Passive galaxies have a significantly higher mean stellar masses despite a similar mass cut (11.07). This is caused by the steeper slope of the stellar mass function of star-forming galaxies and a slight incompleteness of the sample of passive galaxies at low stellar mass (e.g., Ilbert et al. 2013).

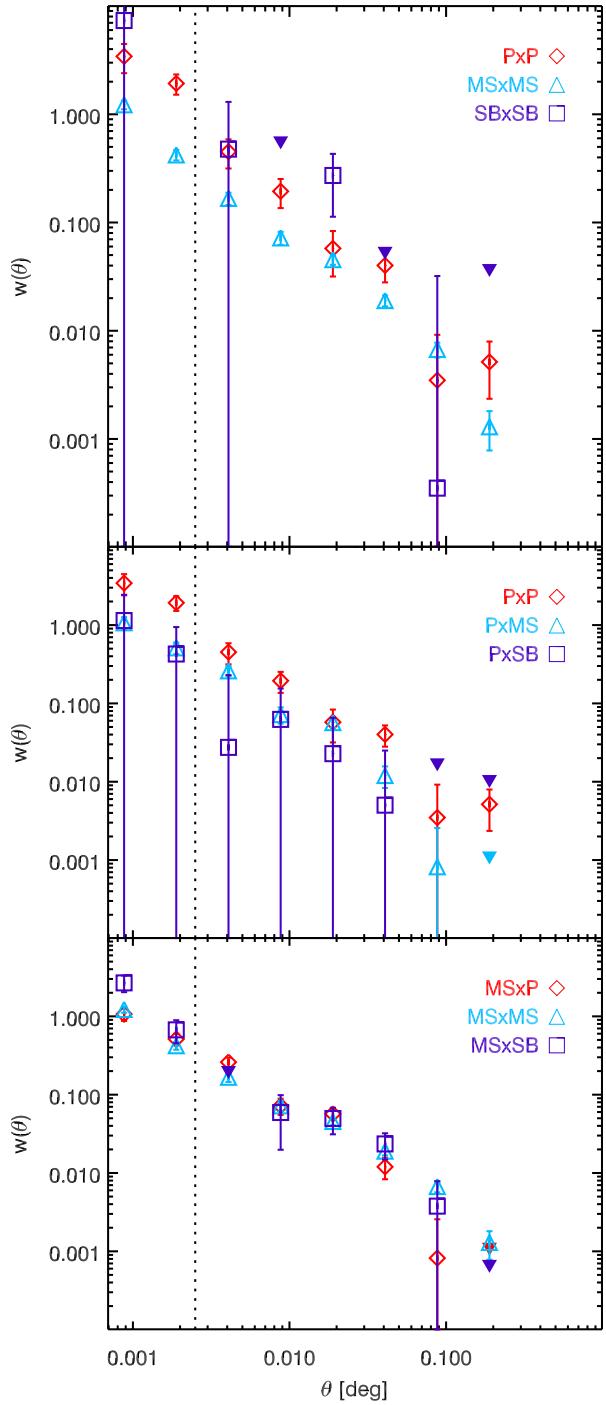

Fig. 10. *Upper panel*: autocorrelation function of the passive (red diamonds), main-sequence (blue triangles), and starburst (purple squares) $\log(M_\star/M_\odot) > 10.5$ galaxies. *Middle panel*: autocorrelation of passive galaxies (red diamonds) compared with the cross-correlation between passive and main-sequence (blue triangles) or starburst (purple squares) galaxies. *Lower panel*: autocorrelation of main-sequence galaxies (blue triangles) compared with the cross-correlation between main-sequence and passive (red triangles) or starburst (purple squares) galaxies. The vertical dotted line indicates the size of the PACS PSF. Below this limit, the reliability of the starburst correlation function is not certain.

The autocorrelation and cross-correlation functions between these populations are shown in Fig. 10. At large scales (>30″), main-sequence and passive galaxies exhibit a similar clustering, but passive galaxies are much more clustered at small scales, as

mentioned previously by McCracken et al. (2010). The origin of this difference is discussed in Sect. 7.3. The autocorrelation signal from the starburst sample is too weak to draw any conclusion. There is no significant difference between the autocorrelation and the cross-correlation functions between the three subsamples at large scale, except a 2.5σ excess of very close ($<10''$) pairs of one starburst and one main sequence galaxy. This could be caused by deblending problems. However, we inspected the 24 μm images visually and in a majority of the case there is a clear separation between the sources in the close pair, suggesting that this excess is not an artifact. These starbursts could be induced because of the interaction with the more massive neighboring main-sequence galaxies, or the interaction might have higher probability to take place in presence of other close neighbors. At smaller scales, all the cross-correlations provide similar results to the autocorrelation of main-sequence galaxies. This suggests that the excess of clustering of passive galaxies at small scales is probably caused by a mechanism affecting only passive galaxies.

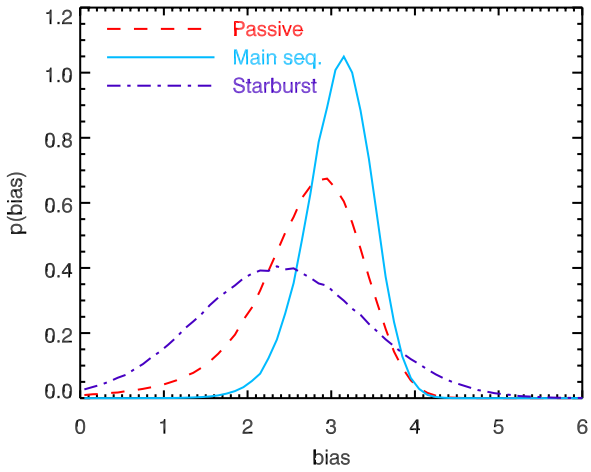

Fig. 11. Probability distribution of the effective bias of passive (red dashed line), main-sequence (blue solid line), and starburst (purple dot-dashed line) galaxies.

7.2. Similar large-scale clustering properties for all three populations

To check if the various types of galaxies (passive, main-sequence, starburst) are hosted by halos with similar halo masses, when they have similar stellar masses, we measured the large-scale bias of these three populations. The bias of passive and main-sequence galaxies can be measured from the ACF only. The bias of the starbursts is much less constrained. For this reason, we also used the cross-correlation functions between these three populations to derive a much better constraint on the bias of starbursts.

Following Eq. (14), we defined the effective bias of a given population by

$$b_{\rm eff} = \int_{M_{\rm h}} \frac{{\rm d}^2N}{{\rm d}\log(M_{\rm h}){\rm d}V} b(M_{\rm h}) \frac{N_{\rm c}+N_{\rm s}}{\bar{n}_{\rm gal}} {\rm d}M_{\rm h}. \quad (19)$$

The two-halo term of the cross power-spectrum between two populations A and B is assumed to be (Cooray & Sheth 2002)

$$P_{\rm A,B} = b_{\rm eff,A}\, b_{\rm eff,B}\, P_{\rm lin}. \quad (20)$$

The two-halo term of the cross-correlation function is thus $\chi_{\rm A,B} = b_{\rm eff,A}\, b_{\rm eff,B}\, \chi_{\rm DM,A,B}$ with (extended version of Eq. (18))

$$\chi_{\rm DM,A,B}(\theta) = \frac{\int_z \frac{{\rm d}N_A}{{\rm d}z}\frac{{\rm d}N_B}{{\rm d}z}\int_k \frac{k}{2\pi}P_{\rm gg}(k,z)J_0(kD_{\rm c}\theta)\,{\rm d}z\,{\rm d}k}{\left(\int_z \frac{{\rm d}N_A}{{\rm d}z}{\rm d}z\right)\left(\int_z \frac{{\rm d}N_B}{{\rm d}z}{\rm d}z\right)}. \quad (21)$$

We thus determined the value of the three effective bias parameters associated with our three subsamples by fitting simultaneously the three autocorrelation functions and the three cross-correlation functions. We used only scales larger than $1'$, where the contribution of the two-halo term is negligible. The confidence regions of these parameters are determined using an MCMC approach.

Figure 11 shows the probability distribution of the effective bias for each population. The effective bias of the three populations are compatible at 1σ: 2.7 ± 0.7 for the passive sample, 3.1 ± 0.4 for the main-sequence sample, and 2.4 ± 0.9 for the starburst sample. This last measurement would have been impossible without this technique based on the cross-correlation, since the constraint provided by the ACF on the bias of strabursts is only an upper limit (<7.0 at 3σ). There is a strong disagreement with Lin et al. (2012) on the bias of passive galaxies (7.1 ± 1.2 in their analysis compared to 2.7 ± 0.7 in our analysis). This means that the bias of passive galaxies at small and large scales is not similar. This result is discussed in Sect. 7.3.

7.3. The origin of the small-scale clustering of passive galaxies

The excess of pairs of passive galaxies at small scales could be explained by the fact that a small fraction of passive galaxies are satellites of central passive galaxies. To test this hypothesis, we selected close pairs of passive galaxies. The fraction of pairs separated by less than $\theta_{\rm max}$, which are associated with galaxies in the same halo (called hereafter intra-halo pairs), can be computed from the autocorrelation function:

$$f_{\rm IH} = \frac{\int_{\theta=0}^{\theta_{\rm max}} w_{\rm 1h}(\theta)\theta\,{\rm d}\theta}{\int_{\theta=0}^{\theta_{\rm max}} [1+w(\theta)]\,\theta\,{\rm d}\theta}. \quad (22)$$

Figure 12 shows this fraction for passive and main-sequence galaxies. This curve was computed from the best HOD fit of the measured ACF of these two populations. We chose to select pairs with a separation smaller than $20''$. This compromise was chosen to have a balance between the small number of pairs found for very small separations and the low purity for large separations. With this cut, we estimate that 58% of the pairs of passive galaxies are intra-halo pairs.

We measured an effective bias squared of all $<20''$ separation pairs of $b^2_{\rm eff,all} = 42 \pm 26$ using only the measurements for $\theta > 1'$, where we can make the linear approximation ($w = b^2_{\rm eff}\, w_{\rm DM}$). We used the barycenter of pairs for our computation of the large-scale clustering. This effective bias of all pairs has to be corrected to estimate only the bias associated with real pairs. The total correlation function can be first-order computed from the biases of the intra-halo pairs and pairs caused by chance alinements ($b_{\rm eff,IH}$ and $b_{\rm eff,CA}$):

$$w = \left[f_{\rm IH} b^2_{\rm eff,IH} + 2\,(1-f_{\rm IH})\, b^2_{\rm eff,CA}\right] w_{\rm DM}. \quad (23)$$

This formula is intuitive if we consider that w is the excess of probability to have a source close to another one compared to the Poisson case. The factor 2 in the second term takes into account

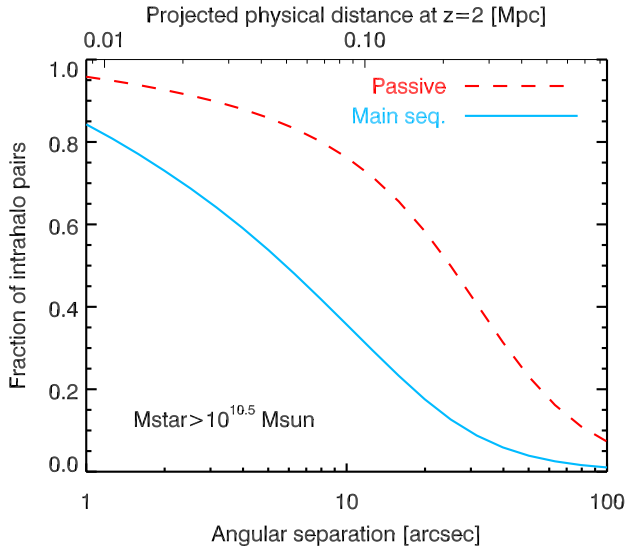

Fig. 12. Fraction of intrahalo pairs as a function of the angular separation for passive (red dashed line) and main-sequence (blue solid line) galaxies.

the excess of probability of finding a source close to both the first and the second sources of pairs caused by chance alignment. This formula neglects the presence of two-halo pairs with similar redshifts, which are negligible for an angular separation lower than 20″. We assume that $b_{\mathrm{eff,CA}}$ is the same as the one of the full population, since only 16% of our population of passive galaxies is a component of a <20″ pair. We found that the bias of intra-halo pairs $b_{\mathrm{eff,IH}}$ is $7.8^{+2.4}_{-3.7}$. This bias corresponds to a halo mass of $5.5^{+5.1}_{-4.5} \times 10^{13}\ M_{\odot}$.

These pairs are thus hosted by structures with a halo mass corresponding to a big group or a small cluster already formed at $z = 2$. This type of structure can potentially have sufficiently massive subhalos to host massive satellite galaxies. We checked that this number agrees with the abundance of such halos. For our assumed cosmology we expect a mean of 17 halos more massive than $5.5 \times 10^{13}\ M_{\odot}$ between $z = 1.4$ and $z = 2.5$ for our field size, compared to ∼80 intracluster pairs based on our estimate of f_{IH}. However, an abundance of 80 halos is reached for a cut of $3 \times 10^{13}\ M_{\odot}$. This value is well inside the 68% confidence region. There is thus no contradiction between the abundances and the clustering. The origin of the excess of clustering at small scales of passive galaxies at $z \sim 2$ is thus probably caused by a small fraction of passive galaxies in massive halos ($\sim 3 \times 10^{13}\ M_{\odot}$).

8. Consistency with X-ray observations

We have found two types of $z \sim 2$ galaxies that appear to be tracing $M_{\mathrm{h}} > 10^{13}\ M_{\odot}$ halos: massive, strongly star-forming, PACS-detected, main-sequence galaxies (see Sect. 5) and close pairs of massive passive galaxies (see Sect. 7.3). At these halo masses, nonnegligible X-ray emissions from hot intra-halo gas might be expected. We thus searched for the X-ray counterparts of our sources to confirm the results of our clustering analysis.

8.1. Direct detections

The COSMOS field has deep X-ray observations by *Chandra* and *XMM-Newton* observatories, allowing us to search for the extended emission down to the level of 8×10^{-16} erg s^{-1} cm^{-2} in the 0.5–2 keV band (Finoguenov et al. 2007; Leauthaud et al. 2010; George et al. 2011). At redshifts above 1, this provides an individual detection of $M_{200} > 5 \times 10^{13}\ M_{\odot}$ groups (as discussed in other high-z COSMOS group papers, e.g., by Onodera et al. 2012). We find 4 close pairs of massive passive galaxies and 13 PACS-detected main-sequence sources are coincident with a directly detected extended X-ray emission and could constitute such sources. Most of the sources are, however, not detected.

8.2. Stacking analysis

We used the emission-free part of the background-subtracted and exposure corrected X-ray image in the 0.5–2 keV band, with the flux of the detected point source removed to make a stacked flux estimate for the undetected sources. We have produced a further background subtraction refinement, by removing the mean residual flux. The PACS-detected, main-sequence sources produce a 4.4σ flux enhancement and pairs produce a marginal 2.2σ flux enhancement at their position, using a 30″ aperture. Accounting for 41% of the pairs being a random association, this corresponds to an average flux of the group of 2.5×10^{-16} erg s^{-1} cm^{-2} for the pairs and 0.9×10^{-16} erg s^{-1} cm^{-2} for the PACS-detected, main-sequence galaxies. Such sources can be individually detected in ultra-deep X-ray surveys, such as CDFS (Finoguenov & et al. 2014), and the corresponding total masses of groups are 3.3 and $2.0 \times 10^{13}\ M_{\odot}$, respectively. These are remarkably similar to the masses inferred from the clustering analysis. The statistical errors on the mean correspond to 0.8 and $0.3 \times 10^{13}\ M_{\odot}$, respectively.

9. Discussion

9.1. How are galaxies quenched at high redshift?

Our measurements indicate that massive ($M_{\star} > 10^{10.5}\ M_{\odot}$), passive galaxies are mostly central galaxies in 10^{12}–$10^{13}\ M_{\odot}$ halos. However, a small fraction of these passive galaxies are satellites in $>10^{13}\ M_{\odot}$ halos as shown in Sect. 7.3. For $M_{\mathrm{h}} < 10^{13}\ M_{\odot}$ halos, where subhalos are not sufficiently massive to host a massive satellite galaxy, the mass quenching is dominant. Our measurements show that this is the dominant process at $z \sim 2$ in agreement with Peng et al. (2010). However, the environmental quenching apparently already had some role at this redshift, as shown by the presence of close pairs of passive galaxies. This role is minor because the number of halos with a sufficiently high mass to host massive galaxies in their substructures is much smaller at $z \sim 2$ than in the local Universe.

We also compared our results with the hydrodynamical simulation of Gabor et al. (2011). In this simulation the main mechanism of quenching is the formation of a hot halo around massive galaxies preventing the accretion of gas and thus the star formation. This hot atmosphere is heated by the winds from supernovae and from the AGN. Figure 13 shows the 3D-correlation function of samples of star-forming and passive galaxies for two various mass cuts ($M_{\star} > 10^{9.5}\ M_{\odot}$ and $M_{\star} > 10^{10.25}\ M_{\odot}$). We used this mass cut of $10^{10.25}\ M_{\odot}$ instead of $10^{10.5}\ M_{\odot}$ because this simulation produces too few massive galaxies at $z \sim 2$ (Gabor & Davé 2012) and we need a sufficient number of objects to have a reasonable signal. In this simulation, the clustering is slightly stronger than measured, because the massive galaxies require more massive host halos to form due to a lack of star formation efficiency in this simulation. There is no difference of clustering in the simulation between $M_{\star} > 10^{10.25}\ M_{\odot}$ star-forming and

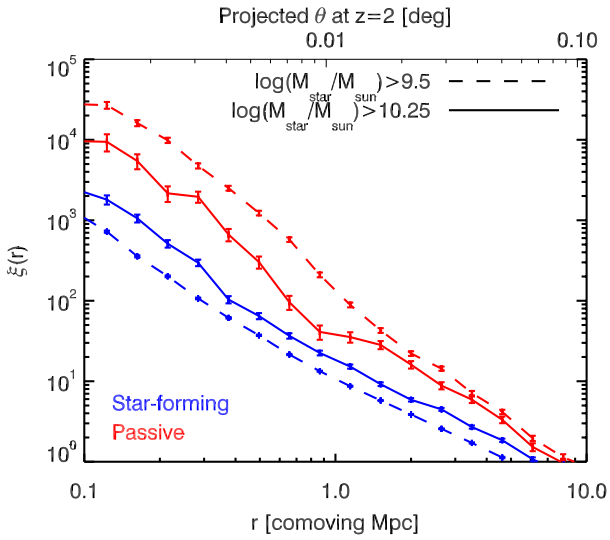

Fig. 13. 3D autocorrelation function of galaxies at $z = 2$ in the hydrodynamical-simulation of Gabor et al. (2011). The results for star-forming and passive galaxies are plotted in blue and red, respectively. The results of two stellar mass cuts are shown: dashed lines for $M_\star > 10^{9.5}\ M_\odot$ and solid lines for for $M_\star > 10^{10.25}\ M_\odot$.

passive galaxies at $r > 0.8$ Mpc (∼30″ at $z = 2$). This is consistent with our clustering measurements. The model also predicts a large excess of clustering for passive galaxies below 0.8 Mpc, which is also observed in the data. This hydrodynamical simulation thus predicts the correct trend, even if the mass assembly is not sufficiently quick. The central galaxies begin to be quenched when they reach a high stellar ($\sim 10^{11}\ M_\odot$) and halo mass ($\sim 10^{12.5}\ M_\odot$). Since the two quantities are correlated, identifying which one is the main driver of the quenching is difficult. The satellite galaxies can be quenched when the hot halo of the central galaxy is formed. This implies that satellite galaxies tends to be quenched more often than centrals at the same mass, and thus the excess of clustering observed at small scales.

Our results are thus compatible with the results from the empirical model of Peng et al. (2010) and the hydrodynamical simulation of Gabor et al. (2011), which both suggests that two different mechanisms of quenching for central and satellite galaxies are already active at $z \sim 2$.

9.2. The future of z ∼ 2 populations

Some interesting insights about galaxy evolution can be obtained by studying not only the instantaneous mass of dark matter halos hosting the $z = 2$ galaxies, but also the mass that these structures will have at $z = 0$. We extrapolate the halo mass at $z = 0$ from the halo mass at $z = 2$ using the mean halo growth of Fakhouri & Ma (2010). Figure 14 is similar to Figs. 5 and 7, but with the halo mass extrapolated at $z = 0$ instead of the instantaneous halo mass. This allows us to connect the $z = 2$ populations studied in our analysis and the descendent population of galaxies at $z = 0$.

The main-sequence galaxies with a stellar mass of a few $10^{10}\ M_\odot$ and a SFR of a few tens of $M_\odot$ per year at $z = 2$ will end up in halos of $M_\mathrm{h} \sim 10^{12.5}$–$10^{13}\ M_\odot$ at $z = 0$. The abundance-matching and weak lensing studies (e.g., Moster et al. 2010; Behroozi et al. 2010; Leauthaud et al. 2012) suggest that these halos host central galaxies with a stellar mass of $\sim 10^{11}\ M_\odot$. At this stellar mass, the mass quenching is efficient according to Peng et al. (2010), and both passive and star-forming galaxies are observed in the low-z Universe (e.g., Baldry et al. 2012; Ilbert et al. 2013). The population of $M_\star \sim 10^{10}\ M_\odot$ sBzK could thus be the progenitor of the most massive field galaxies.

The $M_\star \sim 10^{11}\ M_\odot$ sBzKs and pBzKs, hosted by $M_\mathrm{h} \sim 10^{12.5}\ M_\odot$ $z = 2$ halos, end up in much more massive structures at $z = 0$, with a typical halo mass of $10^{13.5}\ M_\odot$ corresponding to big groups and small clusters of galaxies. A significant fraction of these massive galaxies are already quenched at $z = 2$ and the star-forming ones are probably observed just before their quenching, because the mass function of star-forming galaxies at $M_\star > 10^{11}\ M_\odot$ and especially the massive end does not evolve with redshift as mentioned by Ilbert et al. (2013).

Finally, we identify pairs of passive galaxies (Sect. 7.3) associated with massive structures formed early. An extrapolation of the growth of their host halos at $z = 0$ gives $M_\mathrm{h} = 6.9^{+9.7}_{-6.0} \times 10^{14}\ M_\odot$. These early-formed groups of passive galaxies are thus probably progenitors of the massive clusters in the local Universe, and may be the descendants of the protoclusters of strongly star-forming galaxies observed at $z > 4$ (e.g., Daddi et al. 2009; Capak et al. 2011). This simplified evolution picture is summarized in Fig. 15.

We can also consider the future of SFR-selected population (see Fig. 14 right). The galaxies with $40 < SFR < 200\ M_\odot\,\mathrm{yr}^{-1}$ are progenitors of the central galaxies of groups. The galaxies detected by PACS, if they are episodic starbursts, are expected to be hosted by progenitors of groups in the same way as the rest of the population of sBzK, but PACS-detected main-sequence galaxies are expected to end up in clusters. This suggests that progenitors of clusters can be identified at $z = 2$ using extremely massive star-forming galaxies or groups of massive, passive galaxies. This agrees with Tanaka et al. (2013), who found a diversity of star formation properties of galaxies in X-ray selected groups at $z \sim 1.6$. It is still not clear why some of these structures host star-forming galaxies and others host passive galaxies.

9.3. The nature of the population of starburst galaxies

In the scenario described previously, we neglected the role of the starbursts. As mentioned in the introduction, recent observations favor a scenario where the starbursts host only a minority of the star formation density (∼15%, e.g., Rodighiero et al. 2011; Sargent et al. 2012). However, the mechanisms triggering these violent events are not clearly identified. They could be associated with dense environments such as the group of four SMGs found by Ivison et al. (2013), or those in Daddi et al. (2009) and Chapman et al. (2009). If starbursts are triggered by major mergers, one would naively expect that they are mostly hosted by protoclusters, or some other kind of environmental signature. However, our results show a similar clustering for main-sequence galaxies and starbursts with similar stellar mass, in contradiction with the possibility that starbursts occur in denser environment. In fact, the hydrodynamical simulation of Gabor et al. (2011) shows no excess of clustering above 0.3 Mpc (∼10″ at $z = 2$) for galaxies that merged recently (see Fig. 16). The similar bias thus suggests that $M_\star > 10^{10.5}$ starbursts and main-sequence galaxies are hosted by halos of similar masses. Consequently, starburst episodes do not seem to have a major impact on the $M_\star - M_\mathrm{h}$ relation in agreement with the idea that they have a minor contribution to the star-formation budget. This finding seems to disagree with the claim of Michałowski et al. (2012) that the stellar mass of SMGs is underestimated and that instead they lie on the main sequence, because at fixed SFR they behave as lower mass objects than

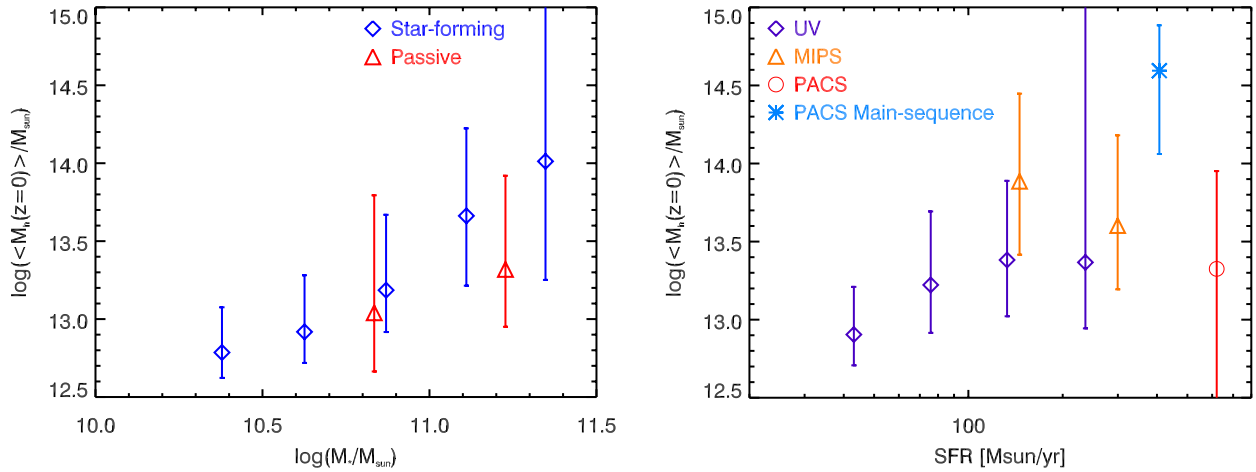

Fig. 14. Mean host halo mass at $z = 0$ (extrapolated from $z = 2$ following Fakhouri & Ma 2010) as a function of the stellar mass (*left panel*) or SFR (*right panel*) at $z \sim 2$. The symbols are similar to those in Figs. 5 and 7.

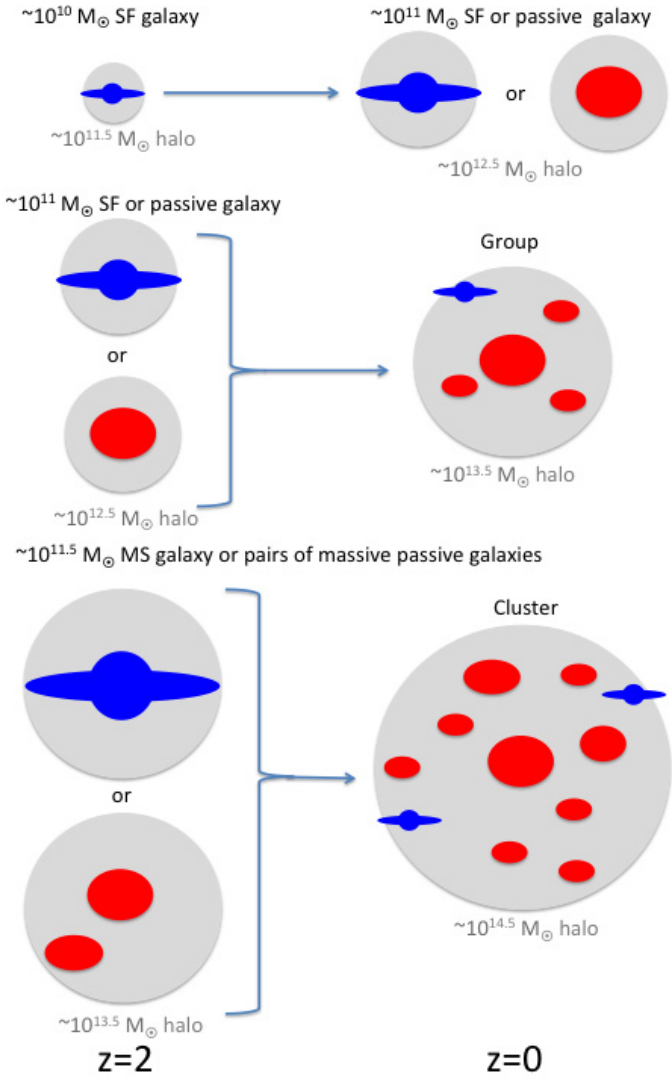

Fig. 15. Simplified evolution scheme from $z = 2$ to $z = 0$ of galaxy populations studied in this paper.

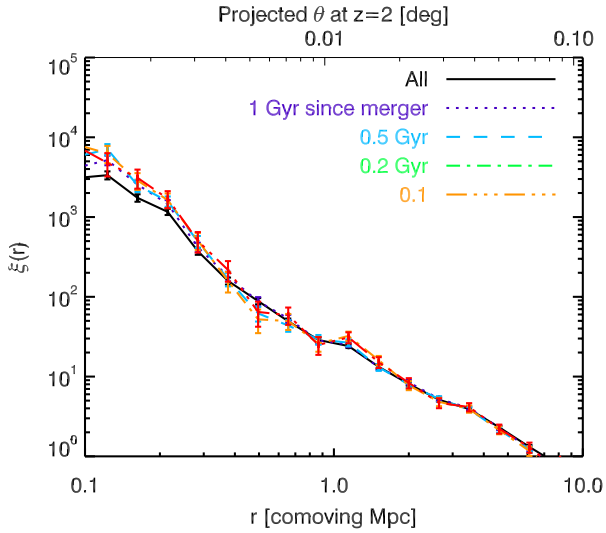

Fig. 16. 3D autocorrelation function of galaxies at $z = 2$ in the Gabor et al. (2011) hydrodynamical simulation as a function of the time since last merger.

main-sequence galaxies. On the other hand, we have found some hints for a possible small-scale cross-correlation excess between starbursts and main-sequence galaxies that could be pointing to some environmental signature. The Gabor et al. (2011) model also predicts such a small-scale enhancement for recent mergers (Fig. 16). This possibility should be further explored in the future.

10. Conclusion

We measured the clustering of a sample of 25 683 star-forming and 2821 passive galaxies in the COSMOS field as a function of their physical properties. This work provided interesting constraints on how the host halo and environment influence the evolution of galaxies at $z \sim 2$, when the cosmic comoving SFR density was at its highest level. Our main findings are as follows:

– We measured the mean host halo mass of $z \sim 2$ passive and star-forming galaxies as a function of their stellar mass using only the clustering. Our results agree well with previous estimates based on abundance matching, suggesting that a monotonic relation with scatter between the stellar and the halo mass is already a fair hypothesis at $z = 2$. We also found similar $M_\star - M_\mathrm{h}$ relations for the two populations.

- We found some hints (2σ) to an increase in the host halo mass with SFR up to 200 $M_\odot\,\mathrm{yr}^{-1}$, where there is a correlation between SFR, $M_\star$ and M_h, and a flattening at higher SFR, where the episodic starbursts with an excess of SFR compared to their stellar mass dominates the population. If we select only main-sequence galaxies, the halo mass continues to rise with SFR above 200 $M_\odot\,\mathrm{yr}^{-1}$. This transition between the main-sequence and the starburst SFR regime happens at the correct position as predicted in the 2SFM model (Sargent et al. 2012; Béthermin et al. 2012a), confirming the relevance of this approach.
- We did not find any difference of large-scale clustering as a function of the sSFR for massive galaxies ($M_\star > 10^{10.5}\ M_\odot$), contrary to Lin et al. (2012) who investigated only smaller scales. This suggests that sSFR does not correlate significantly with the hosting halo mass of the galaxies.
- We confirmed the excess of clustering at small scale ($\theta < 30''$) of passive galaxies found by McCracken et al. (2010). Measuring the large-scale bias of close pairs of passive galaxies, we showed that this is caused by pairs of passive galaxies hosted by the same massive halos ($\sim 3 \times 10^{13}\ M_\odot$). This indicates that the environmental quenching is already operating at $z \sim 2$, even if mass quenching is dominant. This result is in agreement with the empirical model of Peng et al. (2010) and the hydrodynamical simulation of Gabor et al. (2011).
- We finally studied the large-scale bias of the population of starburst galaxies using a method based on the angular cross-correlation function between the various populations of galaxies. We found that the bias of starbursts is similar to the one of main-sequence and passive galaxies of the same stellar mass. This suggests that these three populations live in the halos with similar mass, and that starbursts have only a minor role on the assembly of the stellar mass in the halos. Hints of small scale excess are, however, suggestive of a possible environmental signature.
- Extrapolating the growth of the halos hosting the populations we studied, we predict that the $\sim 10^{10}\ M_\odot$ BzK will end up as massive $\sim 10^{11}\ M_\odot$ field galaxies in the local Universe. The $\sim 10^{11}\ M_\odot$ massive passive and star-forming BzK lie in progenitors of big groups and small clusters. The future massive clusters can be identified by searching for the most massive main-sequence galaxies or groups of massive passive galaxies already formed at $z \sim 2$. The halo mass of these $z = 2$ structures was also confirmed by X-ray stacking.

Thanks to the depth and the area of the COSMOS field from optical to far-infrared, we managed to put first constraints based on clustering measurements on the typical halos hosting the galaxies where the bulk of the star-formation at $z \sim 2$ happen. However, some of the results obtained here are only hints (2σ) or weak evidence (3σ). The deep and very large surveys of the next decade in the optical (e.g., LSST), the near-infrared (e.g., *Euclid*), and the submillimeter (e.g., CCAT) domains will reduce the uncertainties by typically one order of magnitude because of a similar depth to COSMOS, but on fields of $\sim 100\ \mathrm{deg}^2$ or more. This will allow us to study with a much better precision the trends found in this paper and put stronger constraints on the models, but also to have sufficiently large samples to measure the clustering of starbursts at the scale of halos and understand the impact of the environmental effect on them.

Acknowledgements. M.B., E.D., R.G., and V.S. acknowledge the support of the ERC-StG UPGAL 240039 and ANR-08-JCJC-0008 grants. The authors thank Peter Behroozi and Claudia Lagos for providing predictions from their model, and Manuela Magliocchetti and Lihwai Lin for interesting comments.

References

Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, MNRAS, 421, 621
Baugh, C. M., Lacey, C. G., Frenk, C. S., et al. 2005, MNRAS, 356, 1191
Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
Béthermin, M., Daddi, E., Magdis, G., et al. 2012a, ApJ, 757, L23
Béthermin, M., Doré, O., & Lagache, G. 2012b, A&A, 537, L5
Béthermin, M., Le Floc'h, E., Ilbert, O., et al. 2012c, A&A, 542, A58
Béthermin, M., Wang, L., Doré, O., et al. 2013, A&A, 557, A66
Birnboim, Y., Dekel, A., & Neistein, E. 2007, MNRAS, 380, 339
Bouché, N., Dekel, A., Genzel, R., et al. 2010, ApJ, 718, 1001
Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, A&A, 554, A70
Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, Nature, 470, 233
Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J. 2006, MNRAS, 370, 1651
Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, ApJ, 622, 772
Chapman, S. C., Blain, A., Ibata, R., et al. 2009, ApJ, 691, 560
Cimatti, A., Cassata, P., Pozzetti, L., et al. 2008, A&A, 482, 21
Conroy, C., & Wechsler, R. H. 2009, ApJ, 696, 620
Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
Coupon, J., Kilbinger, M., McCracken, H. J., et al. 2012, A&A, 542, A5
Cousin, M., Lagache, G., Blaizot, J., Bethermin, M., & Guiderdoni, B. 2013, A&A, submitted
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Daddi, E., Dannerbauer, H., Stern, D., et al. 2009, ApJ, 694, 1517
Daddi, E., Elbaz, D., Walter, F., et al. 2010, ApJ, 714, L118
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
Fakhouri, O., & Ma, C.-P. 2010, MNRAS, 401, 2245
Finoguenov, A., Tanaka, M., Cooper, M., et al. 2014, A&A, submitted
Finoguenov, A., Guzzo, L., Hasinger, G., et al. 2007, ApJS, 172, 182
Gabor, J. M., & Davé, R. 2012, MNRAS, 427, 1816
Gabor, J. M., Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 417, 2676
George, M. R., Leauthaud, A., Bundy, K., et al. 2011, ApJ, 742, 125
Goldader, J. D., Meurer, G., Heckman, T. M., et al. 2002, ApJ, 568, 651
Gruppioni, C., Pozzi, F., Rodighiero, G., et al. 2013, MNRAS, 432, 23
Guiderdoni, B., Hivon, E., Bouchet, F. R., & Maffei, B. 1998, MNRAS, 295, 877
Hatton, S., Devriendt, J. E. G., Ninin, S., et al. 2003, MNRAS, 343, 75
Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2013, MNRAS, 431, 3373
Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142
Hung, C.-L., Sanders, D. B., Casey, C. M., et al. 2013, ApJ, 778, 129
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, ApJ, 709, 644
Ilbert, O., McCracken, H. J., Le Fevre, O., et al. 2013, A&A, 556, A55
Ivison, R. J., Swinbank, A. M., Smail, I., et al. 2013, ApJ, 772, 137
Kennicutt, Jr., R. C. 1998, ApJ, 498, 541
Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
Lagos, C. D. P., Lacey, C. G., Baugh, C. M., Bower, R. G., & Benson, A. J. 2011, MNRAS, 416, 1566
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16
Le Borgne, D., Elbaz, D., Ocvirk, P., & Pichon, C. 2009, A&A, 504, 727
Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222
Leauthaud, A., Finoguenov, A., Kneib, J.-P., et al. 2010, ApJ, 709, 97
Leauthaud, A., George, M. R., Behroozi, P. S., et al. 2012, ApJ, 746, 95
Lee, K.-S., Giavalisco, M., Conroy, C., et al. 2009, ApJ, 695, 368
Limber, D. N. 1953, ApJ, 117, 134
Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, ApJ, 756, 71
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6
Magliocchetti, M., Santini, P., Rodighiero, G., et al. 2011, MNRAS, 416, 1105
Magliocchetti, M., Lapi, A., Negrello, M., De Zotti, G., & Danese, L. 2014, MNRAS, 437, 2263
Magnelli, B., Popesso, P., Berta, S., et al. 2013, A&A, 553, A132
Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006, MNRAS, 368, 715
McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
Michałowski, M. J., Dunlop, J. S., Cirasuolo, M., et al. 2012, A&A, 541, A85
Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJS, 206, 8
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Onodera, M., Renzini, A., Carollo, M., et al. 2012, ApJ, 755, 26
Peebles, P. J. E. 1980, The large-scale structure of the universe (Princeton University Press)
Peng, Y.-J., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193
Planck Collaboration XXX. 2014, A&A, in press, DOI: 10.1051/0004-6361/201322093
Rodighiero, G., Cimatti, A., Gruppioni, C., et al. 2010, A&A, 518, L25
Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJ, 739, L40
Rodighiero, G., Renzini, A., Daddi, E., et al. 2014, MNRAS, submitted
Salpeter, E. E. 1955, ApJ, 121, 161
Sargent, M. T., Béthermin, M., Daddi, E., & Elbaz, D. 2012, ApJ, 747, L31
Sargent, M. T., Daddi, E., Béthermin, M., et al. 2013, ApJ, submitted [arXiv:1303.4392]
Somerville, R. S., & Primack, J. R. 1999, MNRAS, 310, 1087
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246
Tanaka, M., Finoguenov, A., Mirkazemi, M., et al. 2013, PASJ, 65, 17
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Treister, E., Urry, C. M., Van Duyne, J., et al. 2006, ApJ, 640, 603
Vale, A., & Ostriker, J. P. 2004, MNRAS, 353, 189
van Kampen, E., Percival, W. J., Crawford, M., et al. 2005, MNRAS, 359, 469
Viero, M. P., Wang, L., Zemcov, M., et al. 2013, ApJ, 772, 77
Wang, L., Farrah, D., Oliver, S. J., et al. 2013, MNRAS, 431, 648
Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, MNRAS, 432, 336
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Wuyts, S., Labbé, I., Franx, M., et al. 2007, ApJ, 655, 51
Wuyts, S., Förster Schreiber, N. M., Lutz, D., et al. 2011, ApJ, 738, 106
Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791